

Implementasi *Extra Trees Classifier* dengan Optimasi Grid Search CV pada Prediksi Tingkat Adaptasi

LISTYA NUR AINA, VINNA RAHMAYANTI SETYANING NASTITI, CHRISTIAN SRI KUSUMA ADITYA

Program Studi Informatika, Universitas Muhammadiyah Malang
Email : listyanuraina@webmail.umm.ac.id

Received 9 Januari 2024 | *Revised* 15 Februari 2024 | *Accepted* 7 Maret 2024

ABSTRAK

Teknologi terus maju, terutama dalam komunikasi, pendidikan, dan informasi. Pendidikan online semakin diminati di banyak lembaga pendidikan, mendorong perlunya pemahaman sejauh mana peserta didik dapat beradaptasi dengan lingkungan online. Memprediksi tingkat adaptasi peserta didik menjadi penting untuk meningkatkan efektivitas dan kualitas pengalaman belajar. Dalam penelitian ini, menggunakan dataset dari Kaggle, metode Extra Trees Classifier dioptimalkan dengan Hyperparameter Tuning Grid Search CV. Sebelum optimalisasi, akurasi mencapai 95.85%, setelahnya meningkat menjadi 96.26%, menunjukkan peningkatan sebesar 0.41%. Implementasi metode Extra Trees Classifier dengan optimasi Hyperparameter Tuning Grid Search CV lebih unggul dibandingkan penggunaan algoritma tanpa optimasi.

Kata kunci: *Prediksi, Extra Trees, Classifier, Hyperparameter, CV*

ABSTRACT

Technology continues to advance, especially in communication, education and information. Online education is increasingly in demand in many educational institutions, prompting the need to understand the extent to which learners can adapt to the online environment. Predicting learners' adaptation level is important to improve the effectiveness and quality of the learning experience. In this study, using a dataset from Kaggle, the Extra Trees Classifier method was optimized with Hyperparameter Tuning Grid Search CV. Before optimization, the accuracy reached 95.85%, after which it increased to 96.26%, showing an improvement of 0.41%. The implementation of the Extra Trees Classifier method with Hyperparameter Tuning Grid Search CV optimization is superior to the use of the algorithm without optimization.

Keywords: *Prediction, Extra Trees, Classifier, Hyperparameter, CV*

1. PENDAHULUAN

Pendidikan online atau daring telah berkembang menjadi komponen penting dalam dunia pendidikan modern, terutama seiring dengan kemajuan teknologi informasi dan komunikasi. Pendidikan daring merupakan bentuk pembelajaran yang didukung oleh teknologi yang memudahkan dalam mengakses pendidikan secara luas, berinteraksi dan berpartisipasi tanpa harus bertatap muka. Namun, perubahan dari pembelajaran tatap muka ke pembelajaran daring ini membawa tantangan yang cukup signifikan, terutama dalam mengukur dan memahami tingkat adaptasi siswa terhadap lingkungan pembelajaran daring.

Tingkat adaptasi siswa dalam pendidikan online menjadi penting yang berpengaruh dalam hasil pembelajaran mereka. Siswa yang dapat beradaptasi dengan baik akan mencapai kesuksesan akademik yang lebih baik, sementara yang mengalami kesulitan beradaptasi mungkin memiliki hambatan dalam proses belajar mereka. Salah satu solusi yang dapat digunakan untuk menyelesaikan masalah adaptasi siswa adalah dengan memprediksi tingkat adaptasi mereka pada awal proses pembelajaran. Oleh karena itu, penting untuk memiliki cara yang efektif untuk memprediksi tingkat adaptasi siswa.

Dalam hal ini, *machine learning* menjadi sarana yang efektif dalam memprediksi berbagai konteks termasuk pendidikan (**Jurnal, dkk., 2021**). Pada penelitian (**Divva Meuthia Zulma & Chamidah, 2021**) membandingkan tiga metode klasifikasi, *Decision Trees*, *KNN*, dan *Naives Bayes*, pada data *Log Firewall* pada situs pembelajaran mesin UCI. Setelah tahap pra-pemrosesan, validasi data menggunakan *StratifiedKfold*, dan klasifikasi kumpulan data, *Decision Trees* terbukti memiliki akurasi sebesar 100% dengan nilai AUC sebesar 75% dan kategori AUC yang didapatkan adalah *fair*. Adapun penelitian dari (**Qisthiano, dkk., 2023**) memilih metode *Decision Trees* dan alat bantu *Rapidminer* untuk mengklasifikasi terhadap prediksi kelulusan mahasiswa dari beberapa perguruan tinggi di Kota Palembang. Mereka menggunakan 5 kali proses uji *K-Fold*, hasil akurasi yang didapatkan sebesar 87.93%.

Metode *Extra Trees Classifier* atau dengan nama lainnya *Extremely Randomized Trees Classifier* merupakan salah satu metode dalam *machine learning* yang termasuk metode *ensemble learning* yang digunakan dalam pengenalan aktifitas. Metode ini menggabungkan beberapa *Decision Trees* yang berbeda menjadi sebuah "hutan" dari satu set pembelajaran (*training data*) untuk memprediksi sebuah kejadian atau kelas. Hasil tersebut digunakan untuk memberikan hasil klasifikasi (**Padmaja, dkk., 2020**) (**Rai, dkk., 2022**). Metode *Extra Trees Classifier* merupakan metode yang lebih dikembangkan dan disesuaikan daripada metode *Random Forest*. Berbeda dengan *Random Forest*, *Extra Trees Classifier* menggunakan semua sampelnya untuk membangun model *Decision Trees* dan memilih fitur-fitur yang digunakan secara acak (**Yang & Shami, 2020**).

Pada penelitian yang dilakukan (**Shaaque, dkk., 2019**) yang menganalisis beberapa model klasifikasi seperti *Decision Trees*, *Extra Trees Classifier*, *Random Forest*, *Logistic Regression*, *Support Vector Machine*, *Naives Bayes*. Hasil prediksi dianalisis menggunakan parameter evaluasi seperti *Accuracy*, *Precision*, *Recall*, and *F1-score*. Dan hasilnya menunjukkan akurasi yang paling baik adalah *Extra Trees Classifier* dengan akurasi 90%, sedangkan yang lainnya menghasilkan 88%, 87%, 86%. *Extra Trees Classifier* bekerja jauh lebih cepat daripada *Random Tree Classifier*, bahkan bekerja 3 kali lebih cepat. Adapun juga penelitian dari (**Rustam, dkk., 2021**) yang menganalisis sentimen tweet covid-19. Peneliti menggunakan teknik pra-pemrosesan data dan teknik ekstraksi fitur untuk melatih lima model *machine learning* yaitu *Random Forest*, *XGBoost*, *Support Vector Classifier*, *Extra Trees Classifier*, dan *Decision Treess*. Selain *machine learning*, penelitian ini juga mencoba pendekatan *deep learning* dengan menggunakan arsitektur *Long Short-Term Memory (LSTM)* untuk

menganalisis dataset. Hasilnya menunjukkan *Extra Trees Classifier* mencapai akurasi tertinggi sebesar 0.93, sedangkan LSTM memiliki akurasi lebih rendah dikarenakan ukuran dataset yang kecil.

Kinerja model *machine learning* dapat dioptimalkan dengan menyesuaikan berbagai parameter yang disebut *hyperparameter*. Ada beberapa strategi pencarian yang dapat digunakan untuk menemukan kombinasi *hyperparameter* terbaik dan mampu mengatasi masalah yang diberikan. Cara terbaik untuk menguji berbagai kombinasi parameter dan mengevaluasi kinerja pengklasifikasian yaitu dengan penyesuaian *hyperparameter* (George & Sumathi, 2020).

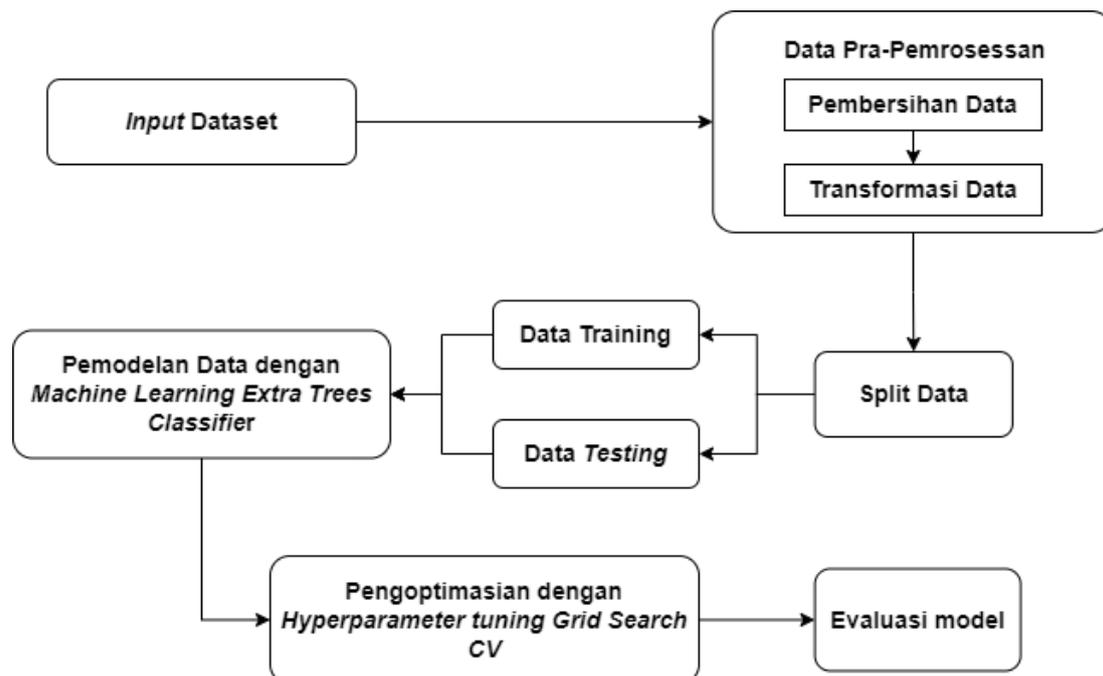
Hyperparameter perlu dikonfigurasi dalam metode *machine learning*, agar lebih baik lagi, kinerja model *machine learning* dioptimalkan dengan menggunakan *hyperparameter tuning* yang efektif untuk dataset tertentu, yang diharapkan mendapatkan hasil yang maksimal. Metode *hyperparameter tuning* yang dapat digunakan antara lain *grid search*, *random search*, *particle swarm optimization*, *Bayesian optimization*, *Genetic Algorithms* dan lain-lain (Elgeldawi, dkk., 2021). Ketika *Grid Search* digunakan untuk mencari kombinasi *hyperparameter* terbaik, setiap kombinasi tersebut akan diuji menggunakan *Cross Validation*. Setiap kombinasi parameter algoritma yang ditetapkan dalam grid dibangun dan dievaluasi oleh *Grid Search Cross Validation (CV)* (George & Sumathi, 2020).

Berdasarkan penelitian sebelumnya yang dilakukan oleh (Mathew, 2022) dengan mengevaluasi kinerja metode klasifikasi *Extra Trees Classifier* untuk mengklasifikasikan kanker payudara. Peneliti membandingkan dengan Metode *Decision Trees* dan lainnya, hasil perbandingan menampilkan bahwa metode *Extra Trees Classifiers* merupakan kinerja yang lebih unggul yaitu dengan akurasi 99.27%. Peneliti juga melakukan optimasi *hyperparameter tuning cross validation* untuk hasil yang lebih optimal. Metode yang digunakan berhasil mengurangi kesalahan dalam mengklasifikasikan pasien kanker payudara, hal ini menunjukkan metode tersebut lebih tepat dan efektif. Adapun juga penelitian dari (Md, dkk., 2023) yang menggunakan 6 algoritma *ensemble learning* untuk mendeteksi penyakit pada hati. Algoritma yang digunakan melakukan berbagai tahapan dalam pemrosesan data dan hasil dianalisa menggunakan berbagai metrik evaluasi. Hasil menunjukkan bahwa *Extra Trees Classifier* memiliki akurasi pengujian tertinggi, yaitu 91.82%, diikuti oleh *Random Forest* dengan akurasi 86.06%. Ketika algoritma dikombinasikan dengan teknik validasi silang 10-fold, akurasi yang dicapai sebesar 93.15%.

Pengimplemantasian metode *Extra Trees Classifier* dan mengoptimalkannya dengan *hyperparameter Tuning Grid Search CV*, diharapkan akan memberikan hasil akurasi prediksi yang lebih akurat. Dari hasil akurasi tersebut dapat memberikan kontribusi dalam meningkatkan kualitas dalam Pendidikan daring, sehingga institusi pendidikan dapat lebih baik lagi dalam mengidentifikasi siswanya yang memerlukan dukungan tambahan dan juga dapat merancang strategi pembelajaran yang lebih efektif dan efisien. Penelitian ini akan mencakup pemahaman yang lebih mendalam tentang bagaimana algoritma *machine learning* dapat diterapkan untuk memahami dan mengukur adaptasi siswa pada pendidikan online.

2. METODOLOGI PENELITIAN

Penelitian ini mengimplementasikan metode *Extra Trees Classifier* dengan optimasi *Hyperparameter Tuning Grid Search CV* untuk mengklasifikasikan tingkat kemampuan adaptasi mahasiswa dalam pendidikan online. Berikut ini adalah langkah-langkah dalam metode penelitian ini :



Gambar 1. Alur Penelitian Model

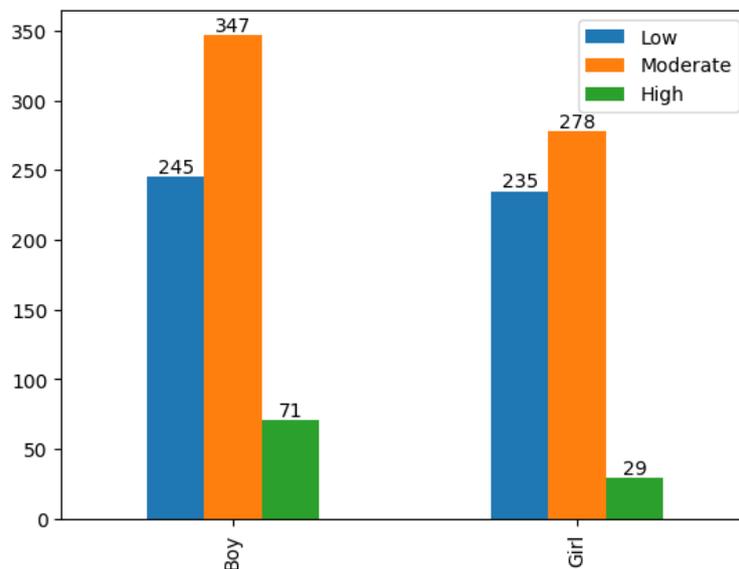
Tahapan penelitian pada Gambar 1. dimulai dari penginputan dataset, lalu dataset tersebut akan melalui pra-pemrosesan seperti pembersihan data dan transformasi data. Setelah melalui pra-pemrosesan, dataset tersebut akan di split menjadi dua yaitu data *training* dan data *testing*. Dataset tersebut akan dimodelkan dengan *Extra Trees Classifier*, dari model tersebut menghasilkan nilai akurasi. Nilai akurasi tersebut akan dioptimasi lagi dengan *Hyperparameter tuning Grid Search CV* yang diharapkan dapat meningkatkan nilai akurasi yang lebih baik. Langkah terakhir yaitu evaluasi model yang digunakan pada penelitian.

2.1. Dataset

Penelitian ini menggunakan data tingkat adaptasi siswa terhadap pendidikan online. Dataset diperoleh dari Kaggle dan berasal dari survei yang dilakukan secara offline dan online pada salah satu institusi sekolah di Bangladesh, India (**Suzan et al., 2021**). Dataset memiliki 1205 data yang terdiri dari 14 atribut, yaitu jenis kelamin, usia, tingkat pendidikan, jenis Institusi, siswa bidang IT, lokasi tinggal di kota, tingkat beban, kondisi keuangan, tipe internet, tipe konektivitas jaringan internet, durasi kelas, status institusi memiliki LMS, perangkat yang digunakan dan tingkat adaptasi. Tabel 1 menjelaskan rincian dari seluruh 14 atribut yang digunakan dan nilai kelas yang. Pemilihan dataset ini didasarkan pada kesesuaian atribut-atribut tersebut untuk digunakan sebagai input dalam proses klasifikasi pada penelitian ini. Dataset ini juga mencakup atribut-atribut yang berkaitan erat dengan tingkat adaptasi mahasiswa dalam pendidikan online. Pada Gambar 2 akan menampilkan visualisasi atribut jenis kelamin dalam bentuk diagram batang.

Tabel 1. Atribut dan Nilai Dataset yang Digunakan

No	Atribut	Nilai
1.	Jenis kelamin	Boy, Girl
2.	Usia	1-5, 6-10, 11-15, 16-20, 21-25, 26-30, 30+
3.	Tingkat pendidikan	School, University, College
4.	Jenis institusi	Non Government, Government
5.	Siswa bidang IT	Yes, No
6.	Lokasi tinggal di kota	Yes, No
7.	Tingkat beban	Low, High
8.	Kondisi Keuangan	Poor, Mid, Rich
9.	Tipe jaringan internet	2G, 3G, 4G
10.	Tipe konektivitas jaringan internet	Tab, Mobile, Computer
11.	Perangkat yang digunakan saat di kelas	Mobile Data, Wifi
12.	Durasi kelas sehari- hari	0, 1-3 Hours, 3-6 Hours
13.	Status institusi memiliki LMS	Yes, No
14.	Tingkat adaptasi	Low, Moderator, High



Gambar 2. Diagram Batang dari Adaptasi Atribut Jenis Kelamin

2.2. Data Pra-Pemrosesan

Sebelum dataset diolah, dilakukan pra-pemrosesan data untuk dapat diklasifikasikan dengan mudah. Pra-pemrosesan data adalah tugas yang penting untuk membersihkan data asli dalam model *machine learning* dan dengan demikian meningkatkan akurasi dan efisiensi model (George & Sumathi, 2020). Pra-pemrosesan data pada penelitian ini meliputi pembersihan data, transformasi data.

Nilai-nilai yang tidak penting atau tidak tersedia (null) dihilangkan dalam proses pembersihan data (Alexandropoulos, dkk., 2019). Pada dataset yang digunakan, proses ini dilakukan terlebih dahulu dengan memeriksa nilai-nilai yang termasuk dalam kategori *missing value* pada setiap atribut. Hasil dari proses tersebut menunjukkan *False* pada setiap atribut yang berarti tidak ada nilai yang hilang dan dianggap bahwa dataset memiliki kualitas yang baik.

Transformasi data dimaksudkan untuk mengubah skala pengukuran dari nilai asli ke dalam bentuk nilai yang lain (Md, dkk., 2023) (Muhamad Malik Matin, 2023). Pada dataset yang digunakan, nilai dari semua atribut merupakan data kategorikal sehingga proses transformasi data dilakukan dengan teknik transformasi menggunakan Standard Scaler, yang pada awalnya nilai data kategorikal diubah menjadi data numerik yang dinormalisasi. Pada Tabel 2 memberikan hasil dari transformasi data, data yang ditampilkan pada table merupakan 5 data teratas dari dataset.

Tabel 2. Dataset yang telah Ditransformasi

Jenis Kelamin	Usia	Tingkat Pendidikan	Jenis Institusi	Siswa Bidang IT	Lokasi Tinggal di Kota	Tingkat Beban	Kondisi Keuangan	Tipe Internet	Tipe Konektivitas	Durasi Kelas	Status LMS	Perangkat	Tingkat Adaptasi
0	3	2	1	0	1	1	0	1	2	2	0	2	2
1	3	2	1	0	1	0	0	0	2	1	1	1	2
1	2	0	0	0	1	1	0	1	2	1	0	1	2
1	1	1	1	0	1	1	0	0	2	1	0	1	2
1	2	1	1	0	1	1	1	0	1	0	0	1	1

Dataset dibagi menjadi dua bagian setelah seluruh pengecekan selesai, dengan perbandingan 8:2 untuk masing-masing data *training* dan data *testing* (Matin, 2023), hasil pembagian tersebut akan tertera pada Tabel 3. Lalu dilanjutkan dengan menentukan objek data dengan masing-masing pembagian terdiri dari data fitur (x) dan data target (y). Data Fitur (x) berisikan semua atribut kecuali atribut tingkat adaptasi, sedangkan Data target (y) hanya berisikan atribut tingkat adaptasi.

Data *training* merupakan dataset yang digunakan untuk melatih dan mengembangkan model-model *machine learning*. Selain itu, pada data training digunakan cross validation untuk meningkatkan validitas model (Gbenga, dkk., 2021).

Tabel 3. Pembagian Data

Keterangan	Presentase (%)	Jumlah Data
Data <i>Training</i>	80	964
Data <i>Testing</i>	20	241
Total	100	1205

2.3. Extra Tree Classifier

Ensemble learning adalah sebuah teknik dalam *Machine learning* di mana beberapa model atau algoritma digabungkan untuk menghasilkan prediksi yang lebih akurat dan stabil. Salah satu metode yang digunakan dalam *Ensemble learning* adalah *Extra Trees Classifier*. *Extra Trees Classifier* adalah metode klasifikasi yang didasarkan pada algoritma *Decision Trees*. Namun, dibandingkan dengan *Decision Trees* biasa, *Extra Trees Classifier* menggunakan pendekatan yang sedikit berbeda dalam pembentukan setiap pohon dalam ensemble. *Extra Trees Classifier* cenderung menghasilkan *Decision Trees* tambahan yang membantu dalam proses voting untuk memberikan prediksi yang lebih baik. Ini membantu mencegah *overfitting*, yaitu kondisi di mana model terlalu sesuai dengan data *training* dan karenanya tidak generalisasi dengan baik untuk data baru (Shaaque, dkk., 2019). *Extra Trees Classifier* merupakan pengklasifikasi yang membangun banyak pohon dan membagi node menggunakan subset acak tanpa substitusi dan node dibagi atas pemisahan acak, bukan partisi terbaik (Ambesange, dkk., 2020).

2.4 Hyperparameter Tuning Grid Search CV

Metode *hyperparameter tuning* yang dapat digunakan antara lain *Grid Search*, *Random Search*, *Particle Swarm Optimization*, *Bayesian Optimization*, *Genetic Algorithm* dan lain-lain. Pada penelitian ini, optimasi model *Machine learning* menggunakan *Grid Search Cross Validation (Grid Search CV)* yang merupakan parameter optimasi model *Machine learning* dengan cara mengkonstruksi dan mengevaluasi setiap kombinasi parameter algoritma yang ditentukan dalam grid dan juga melakukan pencarian secara menyeluruh terhadap semua kombinasi parameter yang ditentukan.

Dengan *Grid Search CV*, peneliti dapat melakukan pemindaian pada sejumlah *hyperparameter* yang diinginkan. Dalam metode ini, sejumlah kombinasi *hyperparameter* akan diterapkan pada model, dan masing-masing kombinasi akan diperiksa untuk performa menggunakan *cross-validation*. Kombinasi *hyperparameter* dengan performa terbaik akan dipilih sebagai *hyperparameter* terbaik untuk model. Penggunaan *Grid Search CV* juga bergantung pada konfigurasi parameter yang dapat disesuaikan dan dapat divariasikan (**Matin, 2023**). Metode statistik yang digunakan untuk mengevaluasi keakuratan model *machine learning* adalah *cross-validation (CV)* (**Elgeldawi, dkk., 2021**) (**Ranjan, dkk., 2019**).

Penggunaan *hyperparameter tuning* dilakukan pada metode *Extra Trees Classifier*. Parameter yang dapat divariasikan dari metode *Hyperparameter Tuning* Menggunakan *Grid Search CV* pada ditunjukkan pada Tabel 4. Hal ini diperlukan agar optimasi dapat berjalan dengan optimal. Pelatihan atau pemasangan model ini membutuhkan 10 kali estimasi waktu.

Tabel 4. Variasi Parameter yang Digunakan

Hyperparameter	Nilai
<i>n_estimators</i>	300
<i>Criterion</i>	<i>entropy</i>
<i>min_samples_split</i>	2
	4
	15
<i>min_weight_fraction_leaf</i>	0.0
	0.15
	0.35
	0.5
<i>max_depth</i>	54

3. HASIL PENELITIAN DAN PEMBAHASAN

Pada tahap ini akan dibahas hasil yang berkaitan dengan metode yang digunakan, termasuk pengolahan data dan evaluasi akhir yang meliputi akurasi, *recall*, *precision*, dan *F1-Score*. Hasil pengujian yang diperoleh sebelum optimasi dan setelah optimasi akan dibandingkan.

3.1. Hasil dari *Extra Trees Classifier*

Hasil evaluasi dari performa model *Extra Trees Classifier* akan ditunjukkan pada Tabel 5. Hasil peningkatan *precision* terus mengalami peningkatan pada adaptabilitas tinggi yaitu sebesar 0.97. *Recall* sebesar 0.95 ini menunjukkan bahwa pada tingkat adaptabilitas rendah dan tinggi dapat mengidentifikasi dari keadaan positif yang sebenarnya. Peningkatan yang terus-menerus dalam *F1-Score* menunjukkan bahwa pada tingkat adaptabilitas tinggi, model *Extra Trees Classifier* memiliki keseimbangan yang sangat baik antara *precision* dan *recall*.

Tabel 5. Hasil Akurasi Menggunakan *Extra Trees Classifier*

<i>Report</i>	<i>Adaptivity Level</i>			<i>Accuracy</i>
	<i>Low</i>	<i>Moderate</i>	<i>High</i>	
<i>Precision</i>	0.86	0.96	0.97	0.95
<i>Recall</i>	0.95	0.93	0.95	
<i>F1-Score</i>	0.90	0.95	0.97	

3.2. Hasil dari *Extra Trees Classifier* dengan *Grid Search CV*

Sedangkan pada Tabel 6 akan membahas mengenai hasil perform model *Extra Trees Classifier* yang sudah dioptimasi dengan *Grid Search CV*. Peningkatan presisi yang signifikan terlihat pada tingkat adaptabilitas sedang pada Tabel 6 setelah dioptimasi. *Recall* pada tingkat adaptabilitas tinggi mengalami peningkatan yang signifikan setelah dioptimasi. Peningkatan *F1-Score* yang signifikan terlihat pada tingkat adaptabilitas rendah dan sedang setelah dioptimasi. Secara keseluruhan, hasil yang dioptimasi menunjukkan kemampuan model yang lebih baik dalam memberikan label positif secara tepat dan mendeteksi *instance* positif yang sebenarnya pada tingkat adaptabilitas yang berbeda.

Tabel 6. Hasil Akurasi yang di Optimasi dengan *Grid Search CV*

<i>Report</i>	<i>Adaptivity Level</i>			<i>Accuracy</i>
	<i>Low</i>	<i>Moderate</i>	<i>High</i>	
<i>Precision</i>	0.86	0.98	0.97	0.96
<i>Recall</i>	0.95	0.93	0.99	
<i>F1-Score</i>	0.97	0.99	0.98	

Berdasarkan hasil pengujian Tabel 5 dan Tabel 6, didapatkan bahwa model *machine learning* dengan algoritma *Extra Trees Classifier* yang dioptimasi menggunakan *hyperparameter tuning Grid Search CV* mendapatkan hasil yang lebih baik dibandingkan dengan algoritma *Extra Trees Classifier* yang tidak dioptimasi. Dapat dilihat dari tabel di atas bahwa akurasi kinerja model sebelum dioptimasi sebesar 95.85%, sedangkan ketika dioptimasi dengan menggunakan metode *hyperparameter tuning Grid Search CV* didapatkan hasil akurasi sebesar 96.26%. Terjadi peningkatan sebesar 0.41% setelah dilakukan optimasi *hyperparameter Tuning Grid Search CV*.

Dengan perbandingan metode yang digunakan pada penelitian (Suzan, dkk., 2021) yang menghasilkan hasil akurasi tertinggi yaitu 89.73% dengan metode *Random Forest Classifier*. Dari hasil penelitian tersebut dapat dibandingkan dengan hasil akurasi metode *Extra Trees Classifier* yang dioptimasi dengan *Hyperparameter Tuning Grid Search CV* yang menghasilkan nilai akurasi 96.26%.

Meskipun dampak yang diberikan tidak signifikan, namun hasil optimasi menggunakan *hyperparameter Tuning Grid Search CV* lebih baik dibandingkan dengan algoritma *Extra Trees Classifier* yang tidak dioptimasi. *Hyperparameter Tuning Grid Search CV* dapat meningkatkan akurasi *Extra trees classifier* karena *Grid Search CV* membantu mencari kombinasi *hyperparameter* terbaik yang dapat meningkatkan kinerja model pada data validasi. Penggunaan "*max_dept*" dan "*min_simple_split*" dapat mencegah *overfitting* pada model yang digunakan.

4. KESIMPULAN

Berdasarkan hasil dan analisis yang telah dilakukan, penelitian ini berhasil mengimplemantasikan metode *Extra Trees Classifier* dengan optimasi *hyperparameter* tuning menggunakan *Grid Search CV* dalam upaya memprediksi tingkat adaptasi siswa dalam konteks Pendidikan online. Dapat diketahui hasil menunjukkan bahwa implementasi metode *Extra Trees Classifier* tanpa optimasi memperoleh hasil akurasi sebesar 95.85%, sedangkan penggunaan metode *Extra Trees Classifier* dengan optimasi *Hyperparameter Tuning Grid Search CV* memperoleh hasil akurasi sebesar 96.26%. Hal ini menunjukkan bahwa *Hyperparameter Tuning Grid Search CV Optimization* mendapatkan hasil yang lebih baik, yang ditunjukkan dengan peningkatan akurasi sebesar 0.41%. Meskipun peningkatan akurasi mungkin tidak signifikan, hasil ini menunjukkan bahwa *hyperparameter tuning* adalah langkah yang relevan dalam meningkatkan kinerja model *machine learning*.

Sebagai saran untuk pengembangan selanjutnya, penelitian ini membuka peluang untuk eksplorasi model *Machine learning* lainnya dalam meningkatkan prediksi tingkat adaptasi siswa. Penelitian ini juga memberi kesempatan untuk pengembangan lebih lanjut, termasuk pengujian algoritma yang lebih beragam guna mencapai performa yang lebih baik.

DAFTAR RUJUKAN

- Alexandropoulos, S. A. N., Kotsiantis, S. B., & Vrahatis, M. N. (2019). Data preprocessing in predictive data mining. *Knowledge Engineering Review*, 34. <https://doi.org/10.1017/S026988891800036X>
- Ambesange, S., Vijayalaxmi, A., Sridevi, S., Venkateswaran, & Yashoda, B. S. (2020). Multiple Heart Diseases Prediction using Logistic Regression with Ensemble and Hyper Parameter tuning Techniques. *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, 827–832. <https://doi.org/10.1109/WorldS450073.2020.9210404>
- Divva Meuthia Zulma, G., & Chamidah, N. (2021). Perbandingan Metode Klasifikasi Naive Bayes, Decision Tree Dan K-Nearest Neighbor Pada Data Log Firewall. In *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA) Jakarta-Indonesia*.
- Elgeldawi, E., Sayed, A., Galal, A. R., & Zaki, A. M. (2021). *Hyperparameter* tuning for machine learning algorithms used for arabic sentiment analysis. *Informatics*, 8(4). <https://doi.org/10.3390/informatics8040079>
- Gbenga, F. O., Olusola, A. A., & Elohor, O. O. (2021). Towards Optimization of Malware Detection using Extra-Tree and Random Forest Feature Selections on Ensemble Classifiers. *The International Journal of Recent Technology and Engineering (IJRTE)*, 9(6), 223–232. <https://doi.org/10.35940/ijrte.F5545.039621>

- George, S., & Sumathi, B. (2020). Grid Search Tuning of *Hyperparameters* in Random Forest Classifier for Customer Feedback Sentiment Prediction. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 11, Issue 9). www.ijacsa.thesai.org
- Jurnal, H., Fathurohman FKIP, A., & Fisika, P. (2021). *JURNAL INFORMATIKA DAN TEKNOLOGI KOMPUTER MACHINE LEARNING UNTUK PENDIDIKAN: MENGAPA DAN BAGAIMANA*. 1(3), 57–62.
- Mathew, T. E. (2022). AN OPTIMIZED EXTREMELY RANDOMIZED TREE MODEL FOR BREAST CANCER CLASSIFICATION. *Journal of Theoretical and Applied Information Technology*, 31(16). www.jatit.org
- Md, A. Q., Kulkarni, S., Joshua, C. J., Vaichole, T., Mohan, S., & Iwendi, C. (2023). Enhanced Preprocessing Approach Using Ensemble Machine Learning Algorithms for Detecting Liver Disease. *Biomedicines*, 11(2). <https://doi.org/10.3390/biomedicines11020581>
- Muhamad Malik Matin, I. (2023). *Hyperparameter* Tuning Menggunakan GridsearchCV pada Random Forest untuk Deteksi Malware. *MULTINETICS*, 9(1), 43–50. <https://doi.org/10.32722/multinetics.v9i1.5578>
- Padmaja, B., Prasad, V. V. R., & Sunitha, K. V. N. (2020). A novel random split point procedure using extremely randomized (Extra) trees ensemble method for human activity recognition. *EAI Endorsed Transactions on Pervasive Health and Technology*, 6(22), 1–10. <https://doi.org/10.4108/eai.28-5-2020.164824>
- Qisthiano, M. R., Prayesy, P. A., & Ruswita, I. (2023). Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa. *G-Tech: Jurnal Teknologi Terapan*, 7(1), 21–28. <https://doi.org/10.33379/gtech.v7i1.1850>
- Rai, N., Kaushik, N., Kumar, D., Raj, C., & Ali, A. (2022). Mortality prediction of COVID-19 patients using soft voting classifier. *International Journal of Cognitive Computing in Engineering*, 3, 172–179. <https://doi.org/10.1016/j.ijcce.2022.09.001>
- Ranjan, G. S. K., Kumar Verma, A., & Radhika, S. (2019, March 1). K-Nearest Neighbors and Grid Search CV Based Real Time Fault Monitoring System for Industries. *2019 IEEE 5th International Conference for Convergence in Technology, I2CT 2019*. <https://doi.org/10.1109/I2CT45611.2019.9033691>
- Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A., & Choi, G. S. (2021). A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *PLoS ONE*, 16(2). <https://doi.org/10.1371/journal.pone.0245909>

- Shaaque, R., Mehmood, A., Choi, G. S., Shafique, R., & Ullah, S. (2019). *Cardiovascular Disease Prediction System Using Extra Trees Classifier Cardiovascular Disease Prediction System Using Extra Trees Classifier*. <https://doi.org/10.21203/rs.2.14454/v1>
- Suzan, M. M. H., Samrin, N. A., Biswas, A. A., & Pramanik, M. A. (2021). Students' Adaptability Level Prediction in Online Education using Machine Learning Approaches. *2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021*. <https://doi.org/10.1109/ICCCNT51525.2021.9579741>
- Yang, L., & Shami, A. (2020). On *hyperparameter* optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, *415*, 295–316. <https://doi.org/10.1016/j.neucom.2020.07.061>