Sistem Klasifikasi ADHD Menggunakan Sinyal EEG Dengan Metode Xgboost

1st Khairurrisya Hanifatuzzahra
Fakultas Teknik Biomedis
Telkom University
Purwokerto, Indonesia
khairurrisya@student.telkomuniversity.ac.id

2nd Mas Aly Afandi
Fakultas Teknik Biomedis
Telkom University
Purwokerto, Indonesia
alyafandi@telkomuniversity.ac.id

3rd Sevia Indah Purnama
Fakultas Teknik Biomedis
Telkom University
Purwokerto, Indonesia
seviaindah@telkomuniversity.ac.id

Abstrak* — *Attention Deficit Hyperactivity Disorder (ADHD) merupakan salah satu gangguan Neurodevelopmental yang dapat memengaruhi perhatian, impulsivitas, dan hiperaktivitas pada anak hingga dewasa. Proses deteksi ADHD saat ini umumnya masih bergantung pada observasi perilaku dan kuesioner yang bersifat subjektif, sehingga diperlukan pendekatan yang lebih objektif. Penelitian ini mengembangkan sistem klasifikasi ADHD berbasis sinyal elektroensefalogram (EEG) menggunakan algoritma Machine learning XGBoost. Data yang digunakan merupakan gabungan dua dataset EEG publik dengan 19 kanal yang diproses melalui tahap Preprocessing dan pembagian data terstratifikasi. Model dilatih menggunakan pendekatan Baseline, optimasi hyperparameter dengan Optuna, serta Cross-validatedTraining untuk meningkatkan kemampuan generalisasi. Hasil eksperimen menunjukkan bahwa model mampu mengklasifikasikan ADHD dan Non-ADHD dengan baik, dengan akurasi sebesar 80,79% dan nilai AUC sebesar 89,53% pada data uji. Hasil ini menunjukkan potensi sistem sebagai alat bantu deteksi ADHD berbasis EEG yang objektif dan berbasis data.

***Kata kunci*— *Attention Deficit Hyperactivity Disorder* (ADHD), *Electroencephalogram* (EEG), XGBoost, Klasifikasi.**

I. PENDAHULUAN

Aktivitas tubuh manusia dikendalikan oleh sistem saraf pusat, khususnya otak, yang berperan dalam perilaku, ingatan, emosi, serta koordinasi fungsi tubuh melalui proses penerimaan dan pengolahan informasi [1]. Gangguan perkembangan sistem saraf (neurodevelopmental disorder) dapat menyebabkan disfungsi struktur dan fungsi otak yang berdampak pada aspek neuropsikologis, salah satunya adalah Attention Deficit Hyperactivity Disorder (ADHD) [2].

ADHD ditandai oleh gejala inatensi, hiperaktivitas, dan impulsivitas yang mengganggu fungsi sehari-hari [2][3]. Penelitian Gehricke Jean-G menunjukkan adanya gangguan pada superior longitudinal fasciculus (SLF) dan sirkuit cortico-limbik yang memengaruhi fungsi fokus, regulasi emosi, memori, dan perilaku melalui keterlibatan amigdala, thalamus, dan korteks prefrontal [4].

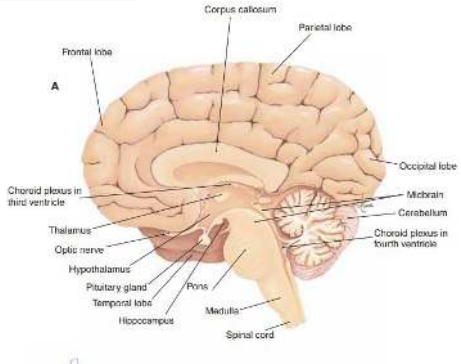
Prevalensi ADHD lebih tinggi pada anak laki-laki dibandingkan perempuan dengan rasio 3–6:1, serta dialami oleh 3%–10% anak usia sekolah, dengan 35%–50% berlanjut hingga remaja dan dewasa [5]. Sebagian besar penderita juga mengalami gangguan komorbid dan memiliki faktor genetik yang kuat, serta berisiko mengalami masalah akademik dan kinerja kerja di masa dewasa [5].

II. KAJIAN TEORI

A. Otak

Otak terletak di dalam tulang tengkorak dan dilindungi oleh lapisan meninges [13]. Rongga tengkorak menampung sekitar 1700 mL yang terdiri dari 80% jaringan otak, 10% darah, dan 10% cairan serebrospinal [14]. Otak manusia tersusun atas sekitar 86 miliar neuron yang memproses informasi dalam bentuk potensial listrik dan berkomunikasi melalui sinapsis [15]. Seluruh bagian otak saling terhubung dan bekerja secara terintegrasi.

Sebagai bagian utama sistem saraf pusat, otak berfungsi mengendalikan dan mengoordinasikan informasi tubuh serta mengatur fungsi kognitif, motorik, emosi, dan homeostasis. Secara anatomi, otak terbagi menjadi serebrum, serebelum, dan batang otak [1]. Serebrum terdiri dari beberapa lobus dengan fungsi spesifik, yaitu lobus frontal (pengambilan keputusan, kontrol impuls, perencanaan, bahasa, dan gerakan), lobus parietal (pemrosesan sensorik), lobus temporal (memori dan pendengaran), serta lobus oksipital (pemrosesan visual). Letak masing-masing lobus ditunjukkan pada Gambar 1 (Anatomi Otak).



GAMBAR 1
(Anatomi Otak [16])

B. *Neurodevelopmental*

Menurut *Diagnostic and Statistical Manual of Mental Disorders* edisi ke-5 (DSM-5), *Neurodevelopmental Disorders* (NDDs) mencakup beberapa kondisi seperti

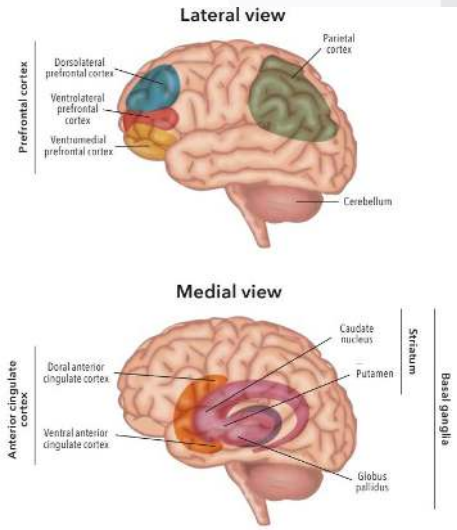
gangguan intelektual, gangguan komunikasi, gangguan spektrum autisme, *attention-deficit hyperactivity disorder* (ADHD), gangguan belajar spesifik, dan gangguan perkembangan motorik. ADHD ditandai dengan gejala inatensi, hiperaktivitas, dan impulsivitas. Gangguan ini dapat memengaruhi kinerja akademik dan hubungan sosial individu. ADHD sering dideteksi pada masa anak-anak, tetapi gejalanya dapat berlanjut hingga dewasa. Gangguan belajar spesifik, seperti disleksia, memengaruhi kemampuan individu dalam membaca, menulis, atau berhitung. Gangguan ini bukan disebabkan oleh kurangnya kecerdasan atau kesempatan belajar, tetapi oleh perbedaan dalam cara otak memproses informasi. Gangguan perkembangan motorik, seperti cerebral palsy, memengaruhi koordinasi dan kontrol gerakan. Gangguan ini dapat disebabkan oleh kerusakan otak yang terjadi sebelum, selama, atau setelah kelahiran.

C. *Attention Deficit Hyperactivity Disorder* (ADHD)

Attention Deficit Hyperactivity Disorder (ADHD) merupakan gangguan neurodevelopmental yang ditandai oleh kesulitan memusatkan perhatian, impulsivitas, dan hiperaktivitas [2]. Penyebab ADHD belum diketahui secara pasti, namun faktor genetik dan lingkungan diketahui berperan dalam meningkatkan risiko terjadinya gangguan ini [18].

Penelitian menunjukkan bahwa anak dengan ADHD memiliki volume otak yang lebih kecil pada beberapa area, termasuk prefrontal cortex, anterior cingulate cortex (ACC), parietal cortex, striatum, dan cerebellum [19]. Studi pencitraan juga menemukan penurunan volume materi abu-abu pada struktur seperti striatum, prefrontal cortex, dan ACC yang berkontribusi terhadap patogenesis ADHD [20].

Struktur otak yang terimplikasi tersebut memiliki fungsi utama yang terganggu pada ADHD, seperti prefrontal cortex dalam perhatian dan pengambilan keputusan, ACC dalam regulasi emosi dan kontrol eksekutif, cerebellum dalam kontrol motorik dan kognitif, parietal cortex dalam orientasi perhatian, serta basal ganglia dalam perilaku berorientasi tujuan, motivasi, dan reward processing [19]. Struktur otak yang terlibat pada ADHD ditunjukkan pada Gambar 2 (Struktur Otak Yang Terimplikasi Penderita ADHD).

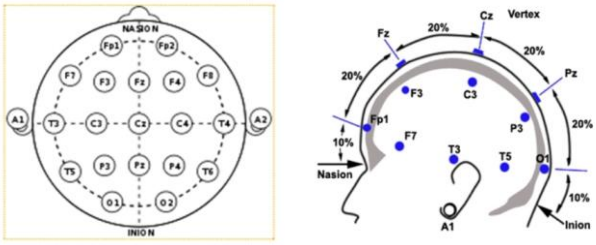


GAMBAR 2
(Struktur Otak Yang Terimplikasi Penderita ADHD [19])

D. Elektroencephalography (EEG)

Elektroensefalografi (EEG) merupakan metode untuk merekam aktivitas potensial listrik otak melalui elektroda yang ditempatkan pada permukaan kulit kepala. Pencatatan sinyal listrik saraf pertama kali dilakukan oleh Carlo Matteucci dan Emil Du Bois-Reymond, yang kemudian diperkuat oleh konsep potensial aksi dari Hermann von Helmholtz. Pada tahun 1875, Richard Caton berhasil merekam aktivitas listrik otak manusia menggunakan galvanometer dengan menempatkan elektroda pada kulit kepala [22]. EEG merekam fluktuasi tegangan yang dihasilkan oleh aktivitas saraf otak secara non-invasif, memungkinkan pemantauan sinyal otak dalam jangka waktu lama [15][23].

Sinyal EEG terdiri dari irama listrik dan pelepasan muatan sementara yang dibedakan berdasarkan lokasi, frekuensi, amplitudo, dan karakteristik fungsionalnya [24]. Penempatan elektroda umumnya mengikuti sistem internasional 10–20 yang direkomendasikan oleh IFSECN, di mana elektroda diletakkan berdasarkan persentase jarak tertentu pada tengkorak [24][25]. Karena EEG rentan terhadap noise, diperlukan tahap praproses seperti bandpass filtering, Independent Component Analysis (ICA), dan spatial filtering untuk menghilangkan gangguan sebelum dilakukan ekstraksi fitur [26].

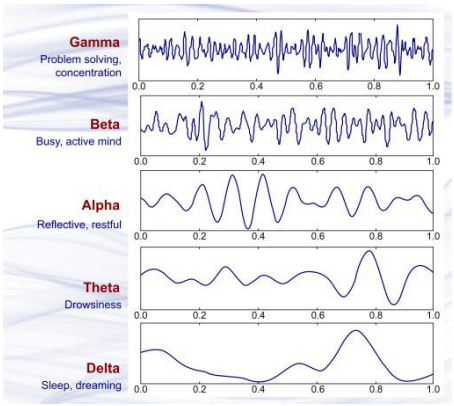


GAMBAR 3
(Penepatan Elektroda Skalp: 10-20 sistem[25])

E. Sinyal EEG

EEG merekam aktivitas listrik otak yang terdiri dari lima gelombang utama, yaitu delta (0,5–4 Hz), teta (4–8 Hz), alfa (8–13 Hz), beta (13–30 Hz), dan gamma (>30 Hz), yang masing-masing berkaitan dengan kondisi dan aktivitas manusia tertentu [15]. Gelombang delta berperan dalam tidur dalam dan regenerasi, teta terkait kondisi mengantuk dan mimpi, alfa berhubungan dengan ketenangan dan relaksasi, beta memengaruhi fokus dan berpikir logis, sedangkan gamma berperan dalam proses kognitif tingkat tinggi [29].

Gelombang beta dibagi menjadi beta rendah (13–15 Hz) yang berkaitan dengan fokus dan konsentrasi, beta tengah (15–20 Hz) yang berhubungan dengan peningkatan energi dan kecemasan, serta beta tinggi (18–35 Hz) yang terkait stres dan energi berlebih [31]. Gelombang alfa berada di antara beta dan teta serta berperan dalam relaksasi dan stabilitas emosi, sedangkan gelombang delta merupakan gelombang paling lambat yang berhubungan dengan tidur restoratif dan sering muncul pada kondisi gangguan otak, termasuk ADHD [31]. Gambar 4(E) menunjukkan sinyal EEG normal..



GAMBAR 4
(Sinyal Normal EEG[31])

F. *Machine learning*

Machine learning (ML) merupakan pengembangan sebuah algoritma dan model yang dapat memungkinkan komputer dapat belajar dari data yang kita berikan[32]. Komputer nantinya akan dilatih untuk mengenali pola dalam membuat sebuah keputusan yang didasari dari data yang akan komputer itu analisis[32]. ML memungkinkan sistem untuk membuat keputusan yang lebih baik bahkan prediksi lebih akurat, sehingga sering dimanfaatkan untuk mengelola data yang banyak dalam berbagai bidang[32].

Machine learning memiliki dua fase yaitu fase pelatihan di mana data dengan label yang sesuai akan dilatih dengan algoritma tertentu sehingga nantinya dapat mengidentifikasi hubungan antara data masukan dan label. Sedangkan fase kedua yaitu fase pengujian, nantinya data tanpa label dimasukkan dengan metodologi yang sama dengan fase pelatihan untuk praproses, ekstraksi fitur, reduksi fitur, dan model yang dilatih, di mana telah diestimasi selama fase pelatihan sehingga memprediksi hasil keluaran. Jadi, fase pelatihan bertujuan untuk memperkirakan model yang memiliki kinerja prediktif maksimal saat pengujian berlangsung[36].

G. Python

Python merupakan bahasa pemrograman tingkat tinggi yang dikembangkan oleh Guido van Rossum dan pertama kali dirilis pada tahun 1991. Bahasa ini dirancang dengan sintaks yang sederhana, mudah dibaca, serta mendukung berbagai paradigma pemrograman seperti prosedural, berorientasi objek, dan fungsional. Python menjadi bahasa yang umum digunakan dalam pengembangan data science dan machine learning karena menyediakan beragam pustaka untuk pemuatan data, visualisasi, statistik, serta pengolahan data secara efisien [30].

Keunggulan utama Python terletak pada struktur kode yang rapi melalui penggunaan indentasi sebagai penentu blok kode, sehingga meningkatkan keterbacaan dan kemudahan pengembangan. Dalam bidang analisis data dan machine learning, Python didukung oleh pustaka seperti pandas dan NumPy untuk manipulasi data numerik, Matplotlib dan Seaborn untuk visualisasi, serta scikit-learn dan XGBoost untuk pengembangan model pembelajaran mesin, menjadikannya fleksibel untuk berbagai aplikasi berbasis AI.

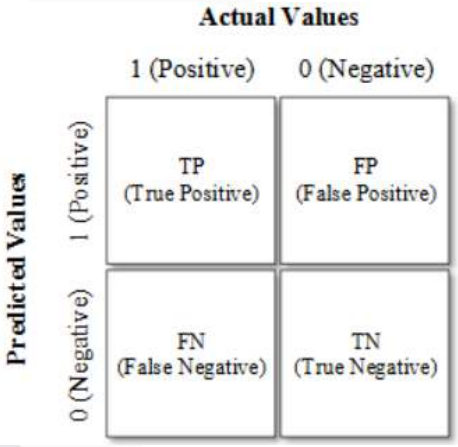
H. XGBoost

Extreme Gradient Boosting (XGBoost) merupakan algoritma *Machine learning* yang dirancang untuk meningkatkan sebuah kinerja dan efisiensi dari algoritma gradient boosting tradisional[38]. XGBoost bekerja dengan membangun model secara interaktif, di setiap iterasinya akan dibangun model baru untuk mengevaluasi kesalahan dari model sebelumnya[39]. XGBoost disebutkan dapat menjalankan tree boosting secara menyeluruh yang dapat diskalakan, metode ini digunakan oleh banyak peneliti yang menggunakan data dalam mencapai sebuah akhir yang mutakhir pada *Machine learning*[39].

XGBoost membuat model baru berdasarkan sisa (error) dari model sebelumnya sehingga dapat memperoleh model baru yang lebih akurat. XGBoost memiliki serangkaian teknik dalam menagani data tidak lengkap, ketidakseimbangan melalui pembobotan, dan kumpulan data besar[39]. Model ini menggunakan teknik boosting yang membangun set model lemah pada subset data dengan meminimalkan nilai mean square error ($\hat{y} - y$) dari model F dengan $\hat{y} = f(x)$ dan menimbang setiap prediksi yang lemah sesuai dengan kinerjanya. ada persamaan[38].

$$\boldsymbol{Obj(t) = \sum_{i}^{n} l\left(\left(y_i \hat{y}_i^{(t-1)} + f_t(x_i)\right) \vdash \Omega(f_t)\right)} \quad (1)$$

Fungsi objektif pada XGBoost ditulis seperti di atas, di mana Ω(f) adalah fungsi regularisasi untuk mencegah *overfitting*. Fungsi ini biasanya mencakup jumlah daun dan besarnya bobot pada setiap daun dalam pohon. Salah satu inovasi utama dari XGBoost adalah penambahan regularisasi L1 (Lasso) dan L2 (Ridge) untuk mengurangi kompleksitas model. Ini membuat XGBoost lebih tahan terhadap *overfitting* dibandingkan algoritma boosting tradisional. XGBoost memiliki fitur bawaan untuk menangani missing.



GAMBAR 5
(*Confusion Matrix* Model[40])

Akurasi merupakan salah satu metrik evaluasi yang paling umum digunakan dalam klasifikasi, termasuk pada model XGBoost. Namun, pada data tidak seimbang *(imbalanced)*, akurasi saja tidak cukup. Metrik lain seperti *precision, recall,* dan *F1-score* juga penting untuk memahami performa model secara menyeluruh. XGBoost secara konsisten menunjukkan akurasi tinggi dibanding algoritma lain seperti *Decision Tree, Logistic Regression,* atau *Naive Bayes*. Akurasi, presisi, *recall, F1-score* dihitung sebagai metrik, yaitu:

$$\boldsymbol{Accuracy = \frac{TP + TN}{TP + TN + FP + FN}} \quad (2)$$

$$\boldsymbol{Precision = \frac{TP}{TP + FP}} \quad (3)$$

$$Recall = \frac{TP}{TP + PN} \quad (4)$$

$$F1 - score = \frac{2xpresisixrecall}{presisixrecall} \quad (5)$$

dimana:

TP (*True Positive*) = jumlah data yang diprediksi dengan benar.
TN *(True Negative*) = jumlah data kontrol yang diprediksi dengan benar.
FP *(False Positive*) = jumlah data kontrol yang salah.
FN *(False Negative)* = jumlah data yang salah diprediksi.

Akurasi model XGBoost sangat dipengaruhi oleh *tuning* parameter seperti *learning_rate, max_depth, n_estimators,* dan *subsample.* Penyesuaian yang tepat dapat meningkatkan akurasi secara signifikan. *Grid* search atau Bayesian optimization sering digunakan untuk menemukan kombinasi parameter terbaik. Dalam konteks deteksi ADHD menggunakan EEG, XGBoost terbukti mampu mencapai akurasi yang tinggi, terutama jika digabungkan dengan teknik ekstraksi fitur seperti band power atau Hjorth parameters. Salah satu tantangan utama dalam mencapai akurasi tinggi adalah *overfitting*, yaitu ketika model sangat akurat pada data latih tetapi buruk pada data uji. XGBoost mengatasi hal ini melalui regularisasi L1 dan L2, sehingga mampu mempertahankan akurasi generalisasi yang baik pada data baru.

I. *Hyperparameter Tuning*

Hyperparameter berperan dalam mengontrol keseimbangan antara kompleksitas model dan kemampuan generalisasi, di mana kompleksitas yang tinggi berisiko menyebabkan overfitting tanpa regularisasi yang memadai. Metode hyperparameter tuning konvensional meliputi manual tuning, grid search, dan random search. Manual tuning bersifat subjektif, grid search memerlukan biaya komputasi tinggi, sedangkan random search lebih efisien namun belum menjamin eksplorasi ruang parameter secara optimal [44].

Optuna merupakan framework optimasi hyperparameter otomatis yang dirancang untuk meningkatkan efisiensi pencarian parameter optimal pada model machine learning. Optuna menerapkan pendekatan define-by-run dengan dua komponen utama, yaitu study sebagai proses optimasi keseluruhan dan trial sebagai satu kali evaluasi fungsi objektif pada kombinasi hyperparameter tertentu. Setiap trial dievaluasi menggunakan metrik performa dan hasilnya dimanfaatkan untuk mengarahkan pencarian parameter berikutnya.

Dalam penelitian ini, metrik evaluasi utama yang digunakan adalah Area Under the Curve (AUC), yang merepresentasikan probabilitas model dalam memberikan skor prediksi lebih tinggi pada sampel positif dibandingkan sampel negatif [42].

$$AUC=P(s(x^+)>s(x^-)) \quad (2.6)$$

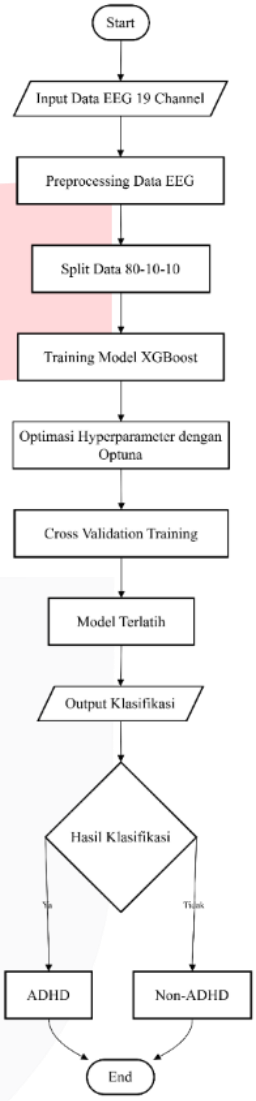
dimana:

s(x^+) = Skor atau probabilitas prediksi yang dihasilkan model untuk sampel dengan kelas positif.
s(x^-) = Skor atau probabilitas prediksi yang dihasilkan model untuk sampel dengan kelas negatif.
P = Menyatakan probabilitas suatu kejadian.

J. METODE

A. Perancangan Sistem

Perancangan sistem penelitian ini menggunakan model XGBoost untuk sistem pengklasifikasian apakah seseorang mengidap ADHD atau tidak dari data sinyal EEG yang diimput. Pada gambar 3 mengilustrasikan tahapan atau proses pengklasifikasian hingga dapat mengeluarkan deteksi.



GAMBAR 5
(Flowchart Sistem)

a. *Start*
Proses dimulai ketika sistem dijalankan dan data EEG telah tersedia.
b. Input Data EEG 19 Channel
Sistem menerima data EEG yang terdiri dari 19 channel sebagai input utama.
c. *Preprocessing* Data EEG
Pada tahap *Preprocessing*, data EEG dipersiapkan agar dapat digunakan secara optimal oleh model *Machine learning*.
d. d. Split Data 80% : 10% : 10%
Data yang telah diproses kemudian dibagi menjadi tiga bagian, yaitu data latih (80%), data validasi (10%), dan data uji (10%).

e. e. *Training* Model XGBoost
Pada tahap ini, model XGBoost dilatih menggunakan data latih.
f. Optimasi Hyperparameter dengan Optuna
Model kemudian melalui proses optimasi hyperparameter menggunakan Optuna.
g. *Cross-Validation Training*
Setelah hyperparameter terbaik diperoleh, model dilatih ulang menggunakan pendekatan cross-validation.
h. Model Terlatih
Hasil dari proses cross-validation menghasilkan model XGBoost yang telah terlatih secara optimal dan stabil.
i. Hasil Klasifikasi
Sistem mengevaluasi hasil klasifikasi dan menentukan apakah data EEG termasuk ke dalam kategori ADHD atau *Non*-ADHD.
j. Jika Hasilnya *Non*-ADHD
Jika hasil klasifikasi menunjukkan *Non*-ADHD, maka sistem menyimpulkan bahwa pola EEG yang dianalisis tidak menunjukkan karakteristik ADHD.
k. Jika Hasilnya ADHD
Jika hasil klasifikasi menunjukkan ADHD, maka sistem menyimpulkan bahwa pola EEG yang dianalisis memiliki karakteristik yang sesuai dengan kelas ADHD.
l. End
Proses klasifikasi selesai dan sistem menghentikan eksekusi.

Dataset pertama merupakan dataset publik yang umum digunakan sebagai benchmark klasifikasi ADHD berbasis EEG, terdiri dari data EEG anak dengan ADHD dan kelompok kontrol sehat. Diagnosis ADHD dilakukan oleh psikiater berdasarkan kriteria DSM-IV, sedangkan kelompok kontrol dipastikan tidak memiliki riwayat gangguan psikiatri, neurologis, maupun perilaku berisiko.

Dataset kedua merupakan dataset publik dari Wrocław University of Science and Technology yang tersedia di RepOD Repository, berisi data EEG partisipan dewasa ADHD dan kontrol sehat dengan proporsi gender seimbang. Diagnosis ADHD dikonfirmasi menggunakan instrumen klinis tervalidasi seperti DIVA dan MOXO test, sementara kelompok kontrol tidak memiliki riwayat gangguan neurodevelopmental.

K. HASIL DAN PEMBAHASAN

A. Hasil Percobaan

Dari percobaan yang dilakukan berdasarkan skenario, didapatkan data berupa output eksekusi: pembacaan dataset, split stratified, *Training Baseline*, *tuning* Optuna, *Training* cross-validation, dan evaluasi ensemble pada *test* set.

1) *Baseline*

Berdasarkan Tabel 1 (Ringkasan olahan data) Ringkasan Olahan Data, dataset gabungan yang digunakan dalam penelitian ini memiliki total 14.959.883 baris data, yang menunjukkan skala data yang sangat besar. Jumlah data ini mencerminkan keberagaman segmen sinyal EEG yang berasal dari dua dataset publik yang berbeda, sehingga memberikan landasan yang kuat bagi model *Machine learning* untuk mempelajari pola yang membedakan kondisi ADHD dan *Non*-ADHD secara lebih komprehensif. Jumlah kolom pada dataset adalah 20 kolom, yang terdiri dari 19 fitur EEG dan 1 label kelas. Fitur EEG merepresentasikan nilai amplitudo sinyal dari masing-masing channel sesuai sistem standar 10–20, sedangkan label kelas digunakan untuk menunjukkan kondisi subjek, yaitu ADHD atau kontrol. Struktur ini sesuai dengan kebutuhan algoritma supervised learning, khususnya XGBoost, yang memerlukan pasangan fitur dan label untuk proses pelatihan.

TABEL 1
(Ringkasan olahan data)

Item	Nilai
Baris dataset gabungan	14,959,883
Kolom (fitur + label)	20 (19 fitur EEG + 1 *Class*)
Proporsi split (train/val/*test*)	80% / 10% / 10%
Ukuran train	11,967,907 baris
Ukuran val	1,495,987 baris
Ukuran *test*	1,495,989 baris
Distribusi label (total)	ADHD = 7,556,069 ; *Control* = 7,403,814

TABEL 2
(Hasil *Baseline* XGBoost)

Metrik	Nilai
Best iteration (early stop)	499
Test Accuracy (Baseline)	0.68357 (68.36%)
Test AUC *(Baseline)*	0.75818
Precision (class 0, Control)	0.67
Recall (class 0, Control)	0.72
Precision (class 1, ADHD*)*	0.70
Recall (class 1, ADHD*)*	0.65
Support (test)	class0: 740,382 class1: 755,607

Berdasarkan Tabel 2 (Hasil *Baseline* XGBoost) Hasil *Baseline* XGBoost, model *Baseline* dilatih menggunakan parameter awal dan menerapkan mekanisme *Early stopping*, di mana iterasi terbaik diperoleh pada iterasi ke-499. *Early stopping* digunakan untuk mencegah *overfitting* dengan menghentikan pelatihan ketika performa model pada data validasi tidak lagi mengalami peningkatan yang signifikan. Dapat dilihat bahwa model telah mampu melakukan klasifikasi dengan tingkat performa tertentu. Hal ini menunjukkan bahwa data EEG yang digunakan mengandung informasi yang relevan untuk membedakan kondisi ADHD dan *Non*-ADHD, meskipun performa yang dihasilkan masih belum optimal. Nilai akurasi pada *Baseline* model memberikan gambaran umum mengenai ketepatan prediksi model secara keseluruhan. Sementara itu, nilai presisi dan *recall* memberikan informasi tambahan terkait kemampuan model dalam mendeteksi masing-masing kelas secara lebih spesifik.

2) *Hyperparameter Tuning*

Hasil percobaan pada tahap *Hyperparameter Tuning* menunjukkan bahwa proses optimasi parameter memberikan peningkatan performa yang signifikan terhadap model XGBoost yang dikembangkan. Pada tahap ini, optimasi dilakukan menggunakan Optuna, yang berfungsi untuk mencari kombinasi hyperparameter terbaik secara sistematis berdasarkan evaluasi performa menggunakan metrik *Area Under the Curve* (AUC) pada proses cross-validation internal. Berdasarkan Tabel 3 (B) Hasil Optuna, diperoleh

nilai Best CV AUC sebesar 0,85557, yang menunjukkan peningkatan yang cukup besar dibandingkan dengan AUC *Baseline* model sebesar 0,75818. Peningkatan ini mengindikasikan bahwa konfigurasi hyperparameter yang dihasilkan mampu meningkatkan kemampuan model dalam membedakan kelas ADHD dan kontrol secara lebih akurat dan stabil. Parameter learning_rate sebesar 0,07848 menunjukkan tingkat pembelajaran model dalam memperbarui bobot pada setiap iterasi. Nilai learning rate yang relatif kecil dipilih untuk memastikan bahwa proses pembelajaran berlangsung secara bertahap dan stabil, sehingga model tidak langsung menyesuaikan diri secara berlebihan terhadap data latih. Pendekatan ini membantu model dalam menangkap pola kompleks pada data EE0020G tanpa menyebabkan ketidakstabilan selama pelatihan.

TABEL 3
(Hasil Optuna)

Parameter	Nilai terbaik
learning_rate	0.07847833100444467
max_depth	12
min_child_weight	27
Subsample	0.9933921990091825
colsample_bytree	0.8711304839589933
reg_alpha	0.047691575405424444
reg_lambda	0.07558293169618242
Best CV AUC *(Optuna)*	0.8555709709815756

Hasil *Hyperparameter Tuning* menunjukkan bahwa model XGBoost yang dioptimalkan mampu meningkatkan performa klasifikasi secara signifikan dengan tetap menjaga stabilitas model. Meskipun nilai *max_depth* yang relatif besar menimbulkan potensi *overfitting*, kombinasi *min_child_weight* dan parameter regularisasi yang diterapkan berhasil mengendalikan kompleksitas model. Oleh karena itu, diperlukan tahap evaluasi lanjutan menggunakan pendekatan *Cross-validatedTraining* untuk memastikan bahwa peningkatan performa yang diperoleh benar-benar mencerminkan kemampuan generalisasi model.

3) *Cross-validated Training* dan *evaluation final*

Tabel 4 (B)hasil evaluasi akhir menyajikan nilai rata-rata metrik performa yang diperoleh dari proses cross-validation. Nilai rata-rata ini digunakan untuk menggambarkan performa model secara lebih stabil dan mengurangi pengaruh variasi data. Berdasarkan hasil evaluasi akhir, model menunjukkan performa yang lebih baik dibandingkan *Baseline* dan tahap *Hyperparameter Tuning* tanpa cross-validation. Hal ini menandakan bahwa model memiliki kemampuan generalisasi yang lebih baik terhadap data yang belum pernah dilihat sebelumnya.

TABEL 4
(Evaluasi Ensemble pada *Test* Set)

Metrik	Nilai
Test Accuracy (ensemble)	0.8078535336824001 (80.79%)
Test AUC *(ensemble)*	0.8953062171300756 (89,53%)
OOF AUC *(internal)*	0.8890938951145682
Confusion matrix (test)	[[607,712, 132,670], [154,779, 600,828]]
Precision (class 0)	0.80
Recall (class 0)	0.82
Precision (class 1)	0.82
Recall (class 1)	0.80
F1-score (both classes)	0.81
Support (test)	*Class* 0: 740,382 ; *Class* 1: 755,607

Berdasarkan Tabel 4 (B) Evaluasi Ensemble pada *Test* Set, model menghasilkan *test* accuracy sebesar 0,80785 (80,79%) dan *test* AUC sebesar 0,89531 (89,53%). Nilai ini menunjukkan peningkatan yang sangat signifikan dibandingkan *Baseline* model, yang sebelumnya hanya mencapai akurasi sekitar 68% dan AUC sebesar 0,758. Peningkatan ini menegaskan bahwa kombinasi antara *Hyperparameter Tuning* dan strategi pelatihan berbasis cross-validation berhasil meningkatkan kemampuan klasifikasi model secara substansial. Nilai Out-of-*Fold* (OOF) AUC sebesar 0,88909 menunjukkan konsistensi internal model selama proses cross-validation. OOF AUC merepresentasikan performa prediksi model pada data yang tidak digunakan dalam pelatihan masing-masing *fold*. Nilai yang mendekati AUC pada data uji menunjukkan bahwa model tidak mengalami *overfitting* yang signifikan dan mampu mempertahankan performa yang stabil di berbagai subset data.

B. Analisis

Pada Tabel 5 (C) perbandingan kinerja model menunjukkan adanya perbedaan yang cukup signifikan antara model *Baseline* XGBoost tunggal dan model ensemble dalam melakukan klasifikasi data ADHD. Kedua model diuji menggunakan dataset yang sama dengan jumlah data uji sebanyak 1.495.989 sampel, sehingga hasil evaluasi yang diperoleh dapat dibandingkan secara langsung dan adil. Perbedaan performa yang muncul sepenuhnya mencerminkan kemampuan masing-masing pendekatan pemodelan dalam mempelajari pola dari data EEG. Berdasarkan nilai akurasi dan AUC, model ensemble memperlihatkan peningkatan kinerja yang jelas dibandingkan model *Baseline*. Akurasi model *Baseline* berada pada kisaran 68%, sedangkan model ensemble mampu mencapai lebih dari 80%. Hal serupa juga terlihat pada nilai AUC, di mana model ensemble menunjukkan kemampuan diskriminasi kelas yang jauh lebih baik.

TABEL 5
(Perbandingan Sebelum dan Sesudah Hyperparameter*Tuning)*

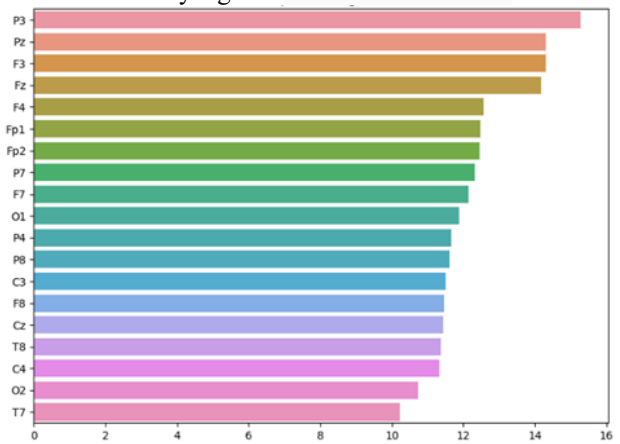
Metrik	***Baseline* Model**	**Ensemble Model**
Test Accuracy	0.6836	0.8079
Test AUC	0.7582	0.8953
Precision – Class 0 (Control)	0.67	0.80
Recall – Class 0 (Control)	0.72	0.82
Precision – Class 1 (ADHD)	0.70	0.82
Recall – Class 1 (ADHD)	0.65	0.80
F1-score (Class 0)	0.69	0.81

F1-score (Class 1)	0.67	0.81
Support Class 0	740,382	740,382
Support Class 1	755,607	755,607
Total Data Uji	1,495,989	1,495,989

Nilai OOF AUC = 0.889 dan Test AUC = 0.895 menunjukkan kemampuan generalisasi model yang kuat, dengan perbedaan kecil (0.006) yang mengindikasikan tidak adanya kebocoran data dan overfitting yang signifikan. Nilai akurasi sebesar 80,8% serta F1-score = 0,81 untuk kedua kelas menunjukkan performa yang seimbang, yang penting untuk aplikasi klinis. Konsistensi metrik pada evaluasi akhir menandakan kestabilan hasil klasifikasi dan memperkuat validitas penelitian melalui penerapan cross-validation.

Hasil evaluasi akhir juga menunjukkan peningkatan yang konsisten dibandingkan tahap sebelumnya, menandakan bahwa model telah mencapai konfigurasi yang optimal. Keseimbangan antara presisi dan recall menunjukkan bahwa model mampu mendeteksi ADHD tanpa mengorbankan klasifikasi Non-ADHD secara signifikan, sehingga model XGBoost layak dipertimbangkan sebagai sistem bantu deteksi ADHD berbasis EEG yang objektif dan efisien.

Gambar 4.2 menunjukkan kontribusi masing-masing channel EEG terhadap kinerja model XGBoost pada fold pertama cross-validation. Sumbu Y menunjukkan channel EEG, sedangkan sumbu X menunjukkan nilai gain sebagai representasi kontribusi fitur dalam meningkatkan performa model. Channel P3, Pz, dan F3 memiliki nilai gain tertinggi, menunjukkan peran paling signifikan dalam membedakan kelas ADHD dan Non-ADHD, sementara channel lain memiliki kontribusi yang lebih kecil.



GAMBAR 6
(Feature Importance)

Adanya kesinambungan antara dasar teori, pengambilan data, dan hasil pemodelan. Seperti pada gambar anatomi otak di Gambar 2 (B) memperlihatkan area mana saja yang terdampak pada penderita ADHD. Adapun sistem elektroda EEG pada Gambar 3 (C) yang menunjukkan bagaimana area-area otak tersebut direpresentasikan dalam bentuk sinyal listrik melalui saluran tertentu. Hasil feature importance dari XGBoost mengungkapkan bahwa model benar-benar menangkap sinyal yang berasal dari area-area otak yang relevan tersebut. Keselarasan ini memperkuat kredibilitas bahwa model tidak hanya belajar pola statistik acak, tetapi pola neurofisiologis yang konsisten dengan teori. Dari pemahaman struktur otak ADHD pada Gambar 2 (B), representasi lokasi sinyal EEG pada Gambar 3 (C), hingga kontribusi fitur pada model XGBoost Gambar 1 (C), menunjukkan hubungan yang kuat antara aspek biologis, pengukuran fisiologis, dan analisis komputasional.

L. KESIMPULAN
M.

Berdasarkan hasil penelitian, sistem klasifikasi ADHD berbasis sinyal EEG menggunakan metode XGBoost berhasil dikembangkan dan dievaluasi secara menyeluruh. Penelitian ini memanfaatkan dataset publik EEG dengan 19 channel yang merepresentasikan berbagai area otak serta pendekatan machine learning berbasis data tanpa melibatkan kuesioner subjektif atau pemeriksaan klinis langsung, sehingga mendukung tujuan pengembangan sistem bantu deteksi ADHD yang lebih objektif.

Hasil eksperimen menunjukkan bahwa model XGBoost yang dioptimasi menggunakan Optuna dan divalidasi dengan K-Fold Cross Validation mampu mempelajari pola EEG dengan baik. Model akhir menghasilkan nilai akurasi sebesar 80,79% dan AUC sebesar 89,53%, dengan nilai precision dan recall yang seimbang pada kedua kelas, yang menandakan kemampuan generalisasi yang baik tanpa bias terhadap kelas ADHD maupun Non-ADHD.

REFERENSI

[1] S. Meutia, N. Utami, S. Rahmawati, dan R. Himayani, “Sistem Saraf Pusat dan Perifer,” 2021.

[2] A. Alhussen, A. I. Alutaibi, S. K. Sharma, A. R. Khan, F. Ahmad, dan G. G. Tejani, “Early attention-deficit/hyperactivity disorder (ADHD) with NeuroDCT-ICA and rhinofish optimization (RFO) algorithm based optimized ADHD-AttentionNet,” Sci Rep, vol. 15, no. 1, Des 2025, doi: 10.1038/s41598-025-90649-1.

[3] Z. Pringgoutami, D. Roro, R. Windi, P. Binaural, R. Rukmi, dan W. Perdani, “Efek Binaural Beats terhadap *Attention Deficit Hyperactivity Disorder* (ADHD),” 2017.

[4] J. G. Gehricke dkk., “The brain anatomy of attention-deficit/hyperactivity disorder in young adults - A magnetic resonance imaging study,” PLoS One, vol. 12, no. 4, Apr 2017, doi: 10.1371/journal.pone.0175433.

[5] A. Budi, “ASPEK NEUROLOGIS *ATTENTION DEFICIT HYPERACTIVITY DISORDER* (ADHD).”

[6] \ Iraary, “DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS FOURTH EDITION DSM-IV TM CHUWEU& MORING.”

[7] H. W. Loh dkk., “ADHD/CD-NET: automated EEG-based characterization of ADHD and CD using explainable deep neural network technique,” Cogn Neurodyn, vol. 18, no. 4, hlm. 1609–1625, Agu 2024, doi: 10.1007/s11571-023-10028-2.

[8] A. Alim dan M. H. Imtiaz, “Automatic Identification of Children with ADHD from EEG Brain Waves,” Signals, vol. 4, no. 1, hlm. 193–205, Mar 2023, doi: 10.3390/signals4010010.

[9] N. Chauhan dan B. J. Choi, “A *Machine learning* Approach to ADHD Diagnosis Using Mutual Information and Stacked Classifiers,” International Journal of Fuzzy Logic and Intelligent Systems, vol. 24,

no. 1, hlm. 10–18, 2024, doi: 10.5391/IJFIS.2024.24.1.10.
[10] M. Cao, E. Martin, dan X. Li, "*Machine learning* in attention-deficit/hyperactivity disorder: new approaches toward understanding the neural mechanisms," 1 Desember 2023, Springer Nature. doi: 10.1038/s41398-023-02536-w.
[11] M. Maniruzzaman, J. Shin, M. A. M. Hasan, dan A. Yasumura, "Efficient Feature Selection and *Machine learning* Based ADHD Detection Using EEG Signal," Computers, Materials and Continua, vol. 72, no. 3, hlm. 5179–5195, 2022, doi: 10.32604/cmc.2022.028339.
[12] N. Fink, "A High-Accuracy Supervised *Machine learning* Approach for ADHD Classification Using EEG Signals." [Daring]. Tersedia pada: https://ssrn.com/abstract=5146511
[13] A. Di Ieva, "BRAIN ANATOMY FROM A CLINICAL AND NEUROSURGICAL PERSPECTIVE A clinically oriented manual of neuroanatomy."
[14] M. S. Amin, "Perbedaan Struktur Otak dan Perilaku Belajar Antara Pria dan Wanita; Eksplanasi dalam Sudut Pandang Neuro Sains dan Filsafat," Jurnal Filsafat Indonesia, vol. 1, no. 1, 2018.
[15] Z. Khakim dan S. Kusrohmaniah, "Dasar - Dasar Electroencephalography (EEG) bagi Riset Psikologi," Buletin Psikologi, vol. 29, no. 1, hlm. 92, Jun 2021, doi: 10.22146/buletinpsikologi.52328.
[16] "-1 Replacing Tissues and Organs," 2007.
[17] F. Shen dan H. Zhou, "Advances in the etiology and neuroimaging of children with attention deficit hyperactivity disorder," 2024, Frontiers Media SA. doi: 10.3389/fped.2024.1400468.
[18] J. Bansal, G. Gangwar, M. Aljaidi, A. Alkoradees, dan G. Singh, "EEG-Based ADHD Classification Using Autoencoder Feature Extraction and ResNet with Double Augmented Attention Mechanism," Brain Sci, vol. 15, no. 1, Jan 2025, doi: 10.3390/brainsci15010095.
[19] P. Curatolo, C. Paloscia, E. D'Agati, R. Moavero, dan A. Pasini, "The neurobiology of attention deficit/hyperactivity disorder," Juli 2009. doi: 10.1016/j.ejpn.2008.06.003.
[20] "The Neurobiology of ADHD," 2020.
[21] P. Song, M. Zha, Q. Yang, Y. Zhang, X. Li, dan I. Rudan, "The prevalence of adult attention-deficit hyperactivity disorder: A global systematic review and meta-analysis," J Glob Health, vol. 11, hlm. 1–9, 2021, doi: 10.7189/jogh.11.04009.
[22] "EEG SIGNAL PROCESSING."
[23] M. Y. Esas dan F. Latifoğlu, "Detection of ADHD from EEG signals using new hybrid decomposition and deep learning techniques," J Neural Eng, vol. 20, no. 3, Jun 2023, doi: 10.1088/1741-2552/acc902.
[24] F. Jurusan Fisika dan F. Matematika dan Ilmu Pengetahuan Alam, "Pengukuran dan Analisa Sinyal Otak Manusia dengan Studi Kasus Pemberian Input Suara Measurement and Analysis of Human Brain Signals with Case Study of Sound Input Provision," Journal of Aceh Physics Society (JAcPS), vol. 6, no. 1, hlm. 10–19, 2017, [Daring]. Tersedia pada: http://www.jurnal.unsyiah.ac.id/JAcPS
[25] "352885-pemeriksaan-EEG-untuk-diagnosis-dan-moni-fb74fa53".
[26] P. L. Nunez dan R. Srinivasan, "Electroencephalography (EEG): neurophysics, experimental methods, and signal processing," 2016, doi: 10.13140/rg.2.2.12706.63687.
[27] Naxon, "Understanding the Strategic Placement of Sensors on EEG devices," Naxon. Diakses: 25 Mei 2025. [Daring]. Tersedia pada: https://naxonlabs-com.translate.goog/blog/understanding-strategic-placement-sensors-EEG-devices?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=sge
[28] M. Adamou, T. Fullen, dan S. L. Jones, "EEG for Diagnosis of Adult ADHD: A Systematic Review With Narrative Analysis," 25 Agustus 2020, Frontiers Media S.A. doi: 10.3389/fpsyt.2020.00871.
[29] J. I. Matematika, R. N. Sofiani, dan D. Juniati, "MATHunesa KLASIFIKASI JENIS EMOSI BERDASARKAN GELOMBANG OTAK MENGGUNAKAN DIMENSI HIGUCHI DENGAN K-NEAREST NEIGHBOR".
[30] J. I. Matematika, R. N. Sofiani, dan D. Juniati, "MATHunesa KLASIFIKASI JENIS EMOSI BERDASARKAN GELOMBANG OTAK MENGGUNAKAN DIMENSI HIGUCHI DENGAN K-NEAREST NEIGHBOR".
[31] P. A. Abhang, B. W. Gawali, dan S. C. Mehrotra, "Technological Basics of EEG Recording and Operation of Apparatus," Introduction to EEG- and Speech-Based Emotion Recognition, hlm. 19–50, Jan 2016, doi: 10.1016/B978-0-12-804490-2.00002-6.
[32] "Pengantar *Machine learning*."
[33] M. Jordan, J. Kleinberg, dan B. Schölkopf, "Pattern Recognition and *Machine learning*."
[34] T. Hastie, R. Tibshirani, dan J. Friedman, "Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction."
[35] A. C. Müller dan S. Guido, "Introduction to *Machine learning* with Python A GUIDE FOR DATA SCIENTISTS Introduction to *Machine learning* with Python."
[36] R. Shoorangiz, S. J. Weddell, dan R. D. Jones, "EEG-Based *Machine learning*: Theory and Applications," dalam Handbook of Neuroengineering, Springer Singapore, 2021, hlm. 1–39. doi: 10.1007/978-981-15-2848-4_70-1.
[37] S. Maesaroh dkk., "Bahasa Pemrograman Python," 2024. [Daring]. Tersedia pada: https://www.researchgate.net/publication/381376588
[38] A. Khairunnisa, "PERBANDINGAN MODEL RANDOM FOREST DAN XGBOOST UNTUK PREDIKSI KEJAHATAN KESUSILAAN DI PROVINSI JAWA BARAT," JIKO (Jurnal Informatika dan Komputer), vol. 7, no. 2, hlm. 202, Sep 2023, doi: 10.26798/jiko.v7i2.799.
[39] M. Wiens, A. Verone-Boyle, N. Henscheid, J. T. Podichetty, dan J. Burton, "A Tutorial and Use Case Example of the eXtreme Gradient Boosting (XGBoost) Artificial Intelligence Algorithm for Drug Development Applications," Clin Transl Sci, vol. 18, no. 3, hlm. e70172, Mar 2025, doi: 10.1111/cts.70172.
[40] Lukman Priyambodo dkk., "Klasifikasi Kematangan Tanaman Hidroponik Pakcoy Menggunakan Metode

SVM," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 6, no. 1, hlm. 153–160, Feb 2022, doi: 10.29207/resti.v6i1.3828.

[41] Y. Cai, J. Feng, Y. Wang, Y. Ding, Y. Hu, dan H. Fang, "The Optuna–LightGBM–XGBoost Model: A Novel Approach for Estimating Carbon Emissions Based on the Electricity–Carbon Nexus," Applied Sciences (Switzerland), vol. 14, no. 11, Jun 2024, doi: 10.3390/app14114632.

[42] "The Effect of SMOTE and Optuna Hyperparameter Optimization on TabNet Performance for Heart Disease Classification".

[43] A. A. TurkaeMLi dan I. D. Sumitra, "A Hyperparameter Optimization Framework Using Optuna for XGBoost-based Drought Zone Classification in Flores Island," 2025, hlm. 64–76. doi: 10.2991/978-94-6463-924-7_7.

[44] A. Tikaningsih, P. Lestari, A. Nurhopipah, I. Tahyudin, E. Winarto, dan N. Hassa, "Optuna Based *Hyperparameter Tuning* for Improving the Performance Prediction Mortality and Hospital Length of Stay for Stroke Patients," Telematika, vol. 17, no. 1, hlm. 1–16, Feb 2024, doi: 10.35671/telematika.v17i1.2816.