ANALYSIS OF GRADIENT BOOSTING ALGORITHMS WITH OPTUNA OPTIMIZATION AND SHAP INTERPRETATION FOR PHISHING WEBSITE DETECTION

LAPORAN NON-REGULER SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika

Disusun oleh :

RAHMAT FAUZI ABU BAKAR

22.11.4857

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2026

ANALYSIS OF GRADIENT BOOSTING ALGORITHMS WITH OPTUNA OPTIMIZATION AND SHAP INTERPRETATION FOR PHISHING WEBSITE DETECTION

LAPORAN NON-REGULER SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika

Disusun oleh :

RAHMAT FAUZI ABU BAKAR

22.11.4857

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2026

HALAMAN PERSETUJUAN

JALUR NON-REGULER SCIENTIST

ANALYSIS OF GRADIENT BOOSTING ALGORITHMS WITH OPTUNA OPTIMIZATION AND SHAP INTERPRETATION FOR PHISHING WEBSITE DETECTION

yang disusun dan diajukan oleh

RAHMAT FAUZI ABU BAKAR

22.11.4857

telah disetujui oleh Dosen Pembimbing

pada tanggal 24 Februari 2026

Dosen Pembimbing,

Mujid Rahardi, S.Kom., M.Eng
NIK. 190302393

HALAMAN PENGESAHAN

JALUR NON-REGULER SCIENTIST

ANALYSIS OF GRADIENT BOOSTING ALGORITHMS WITH OPTUNA OPTIMIZATION AND SHAP INTERPRETATION FOR PHISHING WEBSITE DETECTION

yang disusun dan diajukan oleh

RAHMAT FAUZI ABU BAKAR
22.11.4857

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 Februari 2026

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Arifiyanto Hadinegoro, S.Kom., M.T. **NIK. 190302289**	
Ferian Fauzi Abdulloh, S.Kom., M.Kom. **NIK. 190302276**	
Majid Rahardi, S.Kom., M.Eng. **NIK. 190302393**	

Laporan ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER

Prof. Dr. Kusrini, S.Kom., M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : **RAHMAT FAUZI ABU BAKAR**

NIM : **22.11.4857**

Menyatakan bahwa Laporan dengan judul berikut:

ANALYSIS OF GRADIENT BOOSTING ALGORITHMS WITH OPTUNA OPTIMIZATION AND SHAP INTERPRETATION FOR PHISHING WEBSITE DETECTION

Dosen Pembimbing : Majid Rahardi, S.kom., M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan kegiatan SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidak-benaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 24 Februari 2026

Yang Menyatakan,

Rahmat Fauzi Abu Bakar

HALAMAN PERSEMBAHAN

Karya tulis ini saya persembahkan sebagai wujud rasa syukur dan terima kasih kepada:

1. Allah SWT, atas segala limpahan rahmat, hidayah, dan kekuatan yang diberikan sehingga saya dapat menyelesaikan laporan jalur Scientist ini dengan baik.
2. Kedua Orang Tua tercinta, yang senantiasa memberikan doa tulus, kasih sayang tak terhingga, serta dukungan moril maupun materiil yang menjadi motivasi terbesar saya dalam menuntut ilmu.
3. Bapak Majid Rahardi selaku dosen pembimbing dan *co-author*, yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan serta arahan berharga selama proses penelitian hingga publikasi jurnal.
4. Universitas Amikom Yogyakarta, almamater kebanggaan tempat saya menimba ilmu dan mengembangkan diri.
5. Teman-teman seperjuangan di Program Studi Informatika yang telah memberikan semangat dan dukungan selama masa perkuliahan.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat-Nya, berkat rahmat dan ridha nya yang telah membantu dalam menelesaikan karya ilmiah dengan dengan judul "Analysis of Gradient Boosting Algorithms with Optuna Optimization and SHAP Interpretation for Phishing Website Detection" dapat terselesaikan dengan baik. Penelitian ini disusun untuk menyelesaikan kewajiban dalam memenuhi persyaratan kelulusan program studi Informatika fakultas ilmu komputer Univeristas Amikom Yogyakarta. Penyusunan karya ilmiah ini dapat terselesaikan berkat doa, dukungan, bantuan, serta bimbingan dari berbagai pihak. Pada kesempatan ini, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Prof. Dr. Kusrini, M.Kom., sebagai Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Eli Pujiastuti, M.Kom., sebagai Ketua Program Studi S1 Informatika Universitas AMIKOM Yogyakarta.
4. Bapak Majid Rahardi, S.Kom., M.Eng., selaku Dosen Pembimbing yang telah memberikan arahan berharga dari penyusunan manuskrip hingga laporan ini selesai.
5. Segenap Dosen dan Staf Universitas Amikom Yogyakarta, serta orang tua dan teman-teman yang senantiasa memberikan dukungan doa dan semangat.

Penulis menyadari bahwa laporan ini masih memiliki kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan.

Yogyakarta, 24 Februari 2026

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMBANG DAN SINGKATAN

AI	*Artificial Intelligence* (Kecerdasan Buatan)
APWG	*Anti-Phishing Working Group*
ROC	*Receiver Operating Characteristic*
AUC	*Area Under the Curve (Area di bawah kurva ROC)*
SOC	*Security Operations Center* (Pusat Operasi Keamanan)
LoA	*Letter of Acceptance*
GOSS	*Gradient-based One-Side Sampling*
HTTPS	Hypertext Transfer Protocol Secure
SSL	*Secure Sockets Layer* (Protokol keamanan jaringan)
CRISP-DM	Cross-Industry Standard Process for Data Mining
CatBoost	*Categorical Boosting*
LightGBM	*Light Gradient Boosting Machine*
XGBoost	*eXtreme Gradient Boosting*
SHAP	*SHapley Additive exPlanations* (Metode interpretasi model)
LightGBM	*Light Gradient Boosting Machine*
Σ	Notasi penjumlahan (Sigma)

DAFTAR ISTILAH

Black-box	Istilah untuk model *Machine Learning* yang kompleks (seperti *Ensemble* atau *Neural Networks*) di mana proses internal pengambilan keputusannya sulit dipahami atau dijelaskan oleh manusia.
Cross-Validation	Teknik evaluasi model dengan cara membagi dataset menjadi beberapa bagian (lipatan/folds) untuk melatih dan menguji model secara bergantian guna memastikan konsistensi hasil.
Confusion Matrix	Tabel representasi kinerja model klasifikasi yang membandingkan nilai prediksi dengan nilai sebenarnya, terdiri dari *True Positive*, *True Negative*, *False Positive*, dan *False Negative*.
Ensemble Learning	Metode *Machine Learning* yang menggabungkan prediksi dari beberapa model (biasanya *weak learners* seperti pohon keputusan) untuk menghasilkan satu prediksi akhir yang lebih akurat dan stabil.
False Negative	Kesalahan prediksi di mana model mengklasifikasikan situs berbahaya (*phishing*) sebagai situs aman. Dalam keamanan siber, ini adalah jenis kesalahan yang paling berisiko.
False Positive	Kesalahan prediksi di mana model mengklasifikasikan situs aman (*legitimate*) sebagai situs berbahaya (*phishing*), yang dapat menyebabkan gangguan kenyamanan pengguna.
Gradient Boosting	Teknik algoritma yang membangun model prediksi secara bertahap (*stage-wise*), di mana setiap model baru berusaha memperbaiki kesalahan (*residual error*) dari model sebelumnya.
Grid Search	Metode pencarian hiperparameter konvensional yang mencoba seluruh kombinasi kemungkinan nilai parameter dalam ruang pencarian yang ditentukan, namun sering kali tidak efisien secara komputasi.
Optuna	Kerangka kerja perangkat lunak (*framework*) otomatisasi untuk optimasi hiperparameter yang menggunakan pendekatan efisien dalam menelusuri ruang pencarian parameter.

Hyperparameter	Parameter konfigurasi eksternal model yang nilainya harus ditentukan sebelum proses pelatihan dimulai (contoh: *learning rate*, *max depth*) dan tidak dipelajari dari data latih.
Leaf-wise Growth	Strategi pertumbuhan pohon keputusan (digunakan oleh LightGBM) yang memprioritaskan pembelahan pada daun (*leaf*) dengan nilai kerugian (*loss*) terbesar, sehingga konvergensi lebih cepat namun butuh kontrol agar tidak *overfitting*.
Level-wise Growth	Strategi pertumbuhan pohon keputusan (digunakan oleh XGBoost) yang membelah pohon secara mendatar level demi level, menjaga keseimbangan struktur pohon.
Machine Learning	Cabang dari Kecerdasan Buatan (AI) yang memungkinkan sistem untuk belajar dari data dan meningkatkan kinerjanya tanpa perlu diprogram secara eksplisit untuk setiap aturan.
Phishing	Bentuk serangan siber yang menggunakan teknik rekayasa sosial (*social engineering*) melalui situs web palsu untuk memanipulasi korban agar menyerahkan data sensitif.
SHAP (SHapley Additive exPlanations)	Metode interpretasi model berbasis Teori Permainan (*Game Theory*) yang menjelaskan prediksi model dengan menghitung kontribusi marjinal setiap fitur.
Tree-structured Parzen Estimator	Teknik pembelajaran mesin di mana model yang dikembangkan untuk satu tugas digunakan kembali sebagai titik awal untuk model pada tugas kedua.

INTISARI

Phishing tetap menjadi ancaman keamanan siber yang persisten, berkembang pesat untuk menghindari sistem deteksi tradisional berbasis daftar hitam (*blacklist*). Pendekatan *Machine Learning* (ML) menawarkan solusi yang menjanjikan, namun menemukan kesesimbangan optimal antara akurasi deteksi dan interpretabilitas model masih menjadi tantangan. Penelitian ini bertujuan untuk mengevaluasi dan mengoptimalkan kinerja tiga algoritma *Gradient Boosting* mutakhir XGBoost, LightGBM, dan CatBoost untuk deteksi situs web *phishing*. Penelitian ini memanfaatkan dataset *UCI Phishing Websites* yang terdiri dari 11.055 sampel data. Kebaruan (*novelty*) dari studi ini terletak pada implementasi kerangka kerja Optuna dengan *Tree-structured Parzen Estimator* (TPE) untuk optimasi hiperparameter otomatis dan penerapan nilai interaksi SHAP (*Shapley Additive exPlanations*) untuk menginterpretasikan model "kotak hitam" (*black-box*). Hasil eksperimen menunjukkan bahwa model LightGBM, yang dioptimalkan melalui Optuna, mencapai kinerja tertinggi dengan *F1-Score* sebesar 0,9798, mengungguli model *baseline* (0,9713) sebesar 0,87%. Lebih jauh lagi, analisis SHAP mengidentifikasi 'SSLfinal_State' sebagai determinan paling kritis untuk membedakan situs *phishing*. Penelitian ini mengonfirmasi bahwa optimalisasi algoritma *boosting* modern secara signifikan meningkatkan kemampuan deteksi *phishing* sekaligus memberikan transparansi (*explainability*) yang diperlukan bagi analis keamanan siber..

Kata kunci: Gradient Boosting, Machine Learning, Optuna, Phishing Detection, SHAP

ABSTRACT

Phishing remains a persistent cybersecurity threat, evolving rapidly to bypass traditional blacklist-based detection systems. Machine Learning (ML) approaches offer a promising solution, yet finding the optimal balance between detection accuracy and model interpretability remains a challenge. This study aims to evaluate and optimize the performance of three state-of-the-art Gradient Boosting algorithms—XGBoost, LightGBM, and CatBoost—for phishing website detection. The research utilizes the UCI Phishing Websites dataset consisting of 11,055 instances. The novelty of this study lies in the implementation of the Optuna framework with the Tree-structured Parzen Estimator (TPE) for automated hyperparameter optimization and the application of SHAP (Shapley Additive Explanations) interaction values to interpret the "black-box" models. The experimental results demonstrate that the LightGBM model, optimized via Optuna, achieved the highest performance with an F1-Score of 0.9798, outperforming the baseline model (0.9713) by 0.87%. Furthermore, SHAP analysis identified 'SSLfinal_State' as the most critical determinant for distinguishing phishing sites. This study confirms that optimizing modern boosting algorithms significantly enhances phishing detection capabilities while providing necessary explainability for cybersecurity analysts.

Keyword: *Gradient Boosting, Machine Learning, Optuna, Phishing Detection, SHAP*