ANALISIS PERBANDINGAN TEKNIK EXPLAINABLE AI UNTUK MODEL DETEKSI *PHISING*

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika

disusun oleh

DHANU AHSAN FATIH

21.61.0216

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

ANALISIS PERBANDINGAN TEKNIK EXPLAINABLE AI UNTUK MODEL DETEKSI *PHISHING*

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika

disusun oleh

DHANU AHSAN FATIH

21.61.0216

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

HALAMAN PERSETUJUAN

SKRIPSI

ANALISIS PERBANDINGAN TEKNIK EXPLAINABLE AI UNTUK MODEL DETEKSI PHISHING

yang disusun dan diajukan oleh

Dhanu Ahsan Fatih

21.61.0216

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Juli 2025

Dosen Pembimbing,

Wahid Miftahul Ashari, S.Kom., M.T.
NIK. 190302452

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS PERBANDINGAN TEKNIK EXPLAINABLE AI UNTUK MODEL DETEKSI PHISHING

yang disusun dan diajukan oleh

Dhanu Ahsan Fatih

21.61.0216

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Juli 2025

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Theopilus Bayu S., M.Eng. **NIK. 190302375**	
Agung Nugroho, S.Kom., M.Kom **NIK. 190302242**	
Wahid Miftahul Ashari, S.Kom., M.T. **NIK. 190302452**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 Juli 2025

DEKAN FAKULTAS ILMU KOMPUTER

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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Dhanu Ahsan Fatih
NIM : 21.61.0216

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Perbandingan Teknik Explainable AI untuk Model Deteksi Phishing
Dosen Pembimbing : Wahid Miftahul Ashari, S.Kom., M.T.

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 penelitian 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 ketidakbenaran 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, 21 Juli 2025

Yang Menyatakan,

Dhanu Ahsan Fatih

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur, karya ilmiah ini dipersembahkan kepada Allah SWT atas limpahan rahmat, hidayah, dan kekuatan yang diberikan selama proses pembelajaran dan penyusunan skripsi ini. Kepada kedua orang tua tercinta, yang dengan kasih sayang, doa, dan pengorbanannya menjadi sumber semangat tiada henti. Kepada para dosen dan pembimbing, yang telah tulus membagikan ilmu, arahan, serta bimbingan akademik sepanjang masa studi. Juga kepada seluruh pihak yang telah memberikan dukungan, baik secara moral maupun material, sehingga skripsi ini dapat terselesaikan dengan baik. Semoga karya ini dapat memberikan manfaat serta menjadi bentuk kontribusi nyata bagi pengembangan ilmu pengetahuan, khususnya di bidang teknologi informasi dan keamanan siber.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga skripsi berjudul **"Analisis Perbandingan Teknik Explainable AI untuk Model Deteksi Phishing"** ini dapat diselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada bapak Wahid Miftahul Ashari, S.Kom., M.T. selaku dosen pembimbing atas arahan dan bimbingannya, kepada orang tua atas doa dan dukungannya, serta kepada teman-teman dan semua pihak yang telah membantu dalam proses penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan. Oleh karena itu, saran dan kritik yang membangun sangat diharapkan. Semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilmu dan penelitian di bidang keamanan siber.

Yogyakarta, 21 Juli 2025

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

DAFTAR LAMBANG DAN SINGKATAN

XAI	Explainable Artificial Intelligence
LIME	Local Interpretable Model-Agnostic Explanations
SHAP	SHapley Additive exPlanations
GBM	Gradient Boosting Machines
NLP	Natural Language Processing
DL	Deep Learning
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
MLP	Multi-Layer Perceptron
SVM	Support Vector Machines
DNN	Deep Neural Network
EBM	Explainable Boosting Machine
RF	Random Forest
LR	Logistic Regression
ML	*Machine learning*
IDS	Intrusion Detection System

DAFTAR ISTILAH

Black Box	metode pengujian perangkat lunak di mana struktur internal, desain, dan implementasi item yang diuji tidak diketahui oleh penguji
Adversarial Attack	serangan yang bertujuan untuk menipu model *machine learning* (ML) atau kecerdasan buatan (AI) agar menghasilkan prediksi atau tindakan yang salah atau merugikan
Gradient Boosting Algorithm	teknik pembelajaran mesin yang menggabungkan beberapa model lemah (biasanya pohon keputusan) untuk membentuk model yang lebih kuat dan akurat.
Gini Impurity	Mengukur kemungkinan salah klasifikasi sebuah elemen secara acak jika labelnya dipilih berdasarkan distribusi label di node tersebut.
Entropy	Mengukur jumlah ketidakpastian dalam informasi. Metrik ini berasal dari teori informasi Shannon.
Teori Informasi Shannon	Teori ini menjelaskan bagaimana informasi dikodekan, dikirim, dan diukur secara efisien melalui suatu saluran komunikasi.

INTISARI

Phishing merupakan ancaman siber yang dapat mencuri informasi pribadi melalui situs palsu. Kompleksitas serangan *phishing* membuat sistem deteksi sulit menjelaskan alasan di balik setiap keputusan, sehingga menurunkan kepercayaan pengguna. Oleh karena itu, diperlukan pendekatan yang mampu memberikan transparansi dalam proses deteksi.

Penelitian ini membandingkan beberapa teknik *Explainable Artificial Intelligence* (XAI) dalam menjelaskan model deteksi *phishing* berbasis *machine learning*. Model yang diuji meliputi *Logistic Regression*, *Decision Tree*, *XGBoost, LightGBM, dan CatBoost*. Sementara teknik XAI yang digunakan antara lain *SHapley Additive exPlanations* (SHAP), *Local Interpretable Model-Agnostic Explanations* (LIME), dan *Partial Dependence Plot* (PDP). Dataset yang digunakan terdiri dari fitur-fitur situs phishing dan non-phishing yang telah diproses dan diklasifikasi.

Hasil menunjukkan bahwa teknik XAI dapat menjelaskan berapa kontribusi setiap fitur terhadap prediksi model, terutama SHAP dan LIME yang mampu memberikan visualisasi interpretatif. Penelitian ini diharapkan dapat meningkatkan transparansi dan kepercayaan terhadap sistem deteksi *phishing* serta bermanfaat bagi pengembang sistem keamanan dan peneliti di bidang AI dan keamanan siber.

Kata kunci: phishing, explainable AI, SHAP, LIME, PDP.

ABSTRACT

Phishing is a cyber threat that's designed to steal personal information through deceptive websites. The increasing complexity of phishing attacks makes traditional detection systems less transparent, which can reduce user trust. Therefore, an approach that provides interpretability and transparency in detection decisions is essential.

This research presents a comparative analysis of several Explainable Artificial Intelligence (XAI) techniques applied to phishing detection models. The models evaluated include Logistic Regression, Decision Tree, XGBoost, LightGBM, and CatBoost. The XAI methods used in this study are SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-Agnostic Explanations), and PDP (Partial Dependence Plot). The dataset consists of various features extracted from both phishing and legitimate websites, which are processed and classified using selected *machine learning* algorithms.

The results show that XAI techniques, particularly SHAP and LIME, are effective in explaining the influence of each feature on the model's prediction. This research contributes to enhancing the interpretability and transparency of phishing detection systems, and can be useful for security system developers, end users, and researchers in the field of cybersecurity and artificial intelligence.

Keywords: phishing, explainable AI, SHAP, LIME, PDP.