

**MODEL KLASIFIKASI URL PHISHING MENGGUNAKAN  
DECISION TREE DENGAN ADAPTIVE  
DAN GRADIENT BOOSTING**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi S1 Teknik Komputer



disusun oleh

**KRISNA YOGI SAPUTRA**

**22.83.0795**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2026**

**MODEL KLASIFIKASI URL PHISHING MENGGUNAKAN  
DECISION TREE DENGAN ADAPTIVE  
DAN GRADIENT BOOSTING**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi S1 Teknik Komputer



disusun oleh

**KRISNA YOGI SAPUTRA**

**22.83.0795**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2026**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**MODEL KLASIFIKASI URL PHISHING MENGGUNAKAN  
DECISION TREE DENGAN ADAPTIVE  
DAN GRADIENT BOOSTING**

yang disusun dan diajukan oleh

**KRISNA YOGI SAPUTRA**

**22.83.0795**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 25 Februari 2026

**Dosen Pembimbing,**



**Tonny Hidayat, S.Kom., M.Kom., Ph.D.**

**NIK. 190302182**

**HALAMAN PENGESAHAN**  
**SKRIPSI**  
**MODEL KLASIFIKASI URL PHISHING MENGGUNAKAN**  
**DECISION TREE DENGAN ADAPTIVE**  
**DAN GRADIENT BOOSTING**

yang disusun dan diajukan oleh

**KRISNA YOGI SAPUTRA**

22.83.0795

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 25 Februari 2026

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Muhammad Kopravi, S.Kom., M.Eng.**  
NIK. 190302454

**Jeki Kuswanto, S.Kom., M.Kom.**  
NIK. 190302456

**Tonny Hidayat, S.Kom., M.Kom., Ph.D.**  
NIK. 190302182



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 25 Februari 2026

**DEKAN FAKULTAS ILMU KOMPUTER**



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

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

**Nama mahasiswa : Krisna Yogi Saputra**  
**NIM : 22.83.0795**

Menyatakan bahwa Skripsi dengan judul berikut:

Model Klasifikasi URL Phishing Menggunakan Decision Tree Dengan Adaptive dan Gradient Boosting

Dosen Pembimbing : Tonny Hidayat, S.Kom., M.Kom., Ph.D.

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, 25 Februari 2026

Yang Menyatakan,



Krisna Yogi Saputra

## HALAMAN PERSEMBAHAN

Dengan segala kerendahan hati dan rasa syukur yang mendalam, skripsi ini penulis persembahkan kepada:

1. Allah SWT Atas segala rahmat, hidayah, dan kekuatan yang Engkau berikan di setiap detik perjuangan ini. Tanpa izin dan rida-Mu, langkah ini tidak akan pernah sampai pada titik ini. Terima kasih atas segala kemudahan di tengah kesulitan yang penulis hadapi.
2. Kedua Orang Tua Tercinta, Bapak dan Ibu yang merupakan sumber kekuatan utama bagi penulis. Terima kasih atas doa-doa tulus yang senantiasa dipanjatkan di setiap sujud, kasih sayang yang tidak pernah lekang oleh waktu, serta pengorbanan luar biasa baik tenaga, pikiran, maupun materi demi masa depan penulis. Gelar Sarjana ini adalah persembahkan kecil untuk membalas ketulusan kalian.
3. Adik Perempuan Tersayang yang selalu menjadi penghibur di kala lelah dan pemberi semangat melalui keceriaannya. Terima kasih telah menjadi saudara yang luar biasa dan selalu memberikan dukungan moral selama penulis menyelesaikan studi ini. Semoga pencapaian ini menjadi motivasi bagimu untuk meraih mimpi yang lebih tinggi.
4. Keluarga Besar, yang selalu memberikan dukungan, perhatian, dan doa yang tidak terputus. Terima kasih telah menjadi lingkungan yang hangat dan suportif bagi perjalanan akademik penulis.

## KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT atas segala rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Model Klasifikasi URL Phishing Menggunakan Decision Tree dengan Adaptive dan Gradient Boosting” ini dengan baik.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknik Komputer, Universitas Amikom Yogyakarta. Penulisan skripsi ini tentu tidak lepas dari bantuan, dukungan, serta bimbingan dari berbagai pihak. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua tercinta, Bapak dan Ibu, sebagai pilar utama dalam hidup penulis. Terima kasih atas segala doa yang tidak pernah putus, kasih sayang yang tidak terhingga, serta pengorbanan baik moril maupun materil yang telah diberikan.
2. Adik Terima kasih atas semangat dan keceriaan yang selalu diberikan. Semoga keberhasilan ini menjadi motivasi bagimu untuk terus mengejar cita-cita.
3. Bapak Tonny Hidayat, S.Kom., M.Kom., Ph.D. selaku dosen pembimbing yang telah memberikan arahan, bimbingan, serta motivasi selama proses penyusunan skripsi ini.
4. Seluruh rekan-rekan Kelompok Belajar Kartel dan rekan – rekan kelas TK-01 yang tidak dapat penulis Sebutkan satu per satu, yang telah memberikan bantuan dan semangat dalam penyelesaian skripsi ini.

Yogyakarta, 28 Februari 2026

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PERSETUJUAN .....</b>	<b>ii</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>iii</b>
<b>HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....</b>	<b>iv</b>
<b>HALAMAN PERSEMBAHAN .....</b>	<b>v</b>
<b>KATA PENGANTAR.....</b>	<b>vi</b>
<b>DAFTAR ISI.....</b>	<b>vii</b>
<b>DAFTAR TABEL .....</b>	<b>x</b>
<b>DAFTAR GAMBAR.....</b>	<b>xi</b>
<b>DAFTAR LAMBANG DAN SINGKATAN .....</b>	<b>xii</b>
<b>DAFTAR ISTILAH .....</b>	<b>xiii</b>
<b>INTISARI .....</b>	<b>xv</b>
<b>ABSTRACT .....</b>	<b>xvi</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	4
1.6 Sistematika Penulisan .....	4
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>6</b>

2.1	Studi Literatur .....	6
2.2	Dasar Teori.....	13
2.2.1	Phising.....	13
2.2.2	<i>Machine Learning</i> .....	14
2.2.3	Preprocessing .....	15
2.2.4	Decision Tree .....	16
2.2.5	Adaptive Boosting .....	18
2.2.6	Gradient Boosting .....	21
2.2.7	Training.....	23
2.2.8	Validasi .....	23
2.2.9	Testing.....	24
2.2.10	Evaluasi Model .....	24
<b>BAB III METODE PENELITIAN .....</b>		<b>26</b>
3.1	Objek Penelitian.....	26
3.2	Alur Penelitian .....	27
3.3	Skenario Pengujian .....	38
3.4	Analisis Kebutuhan .....	40
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>44</b>
4.1	Persiapan Lingkungan Penelitian.....	44
4.2	Persiapan Library .....	44
4.3	Deskripsi Dataset .....	45
4.4	Tahap Preprocessing .....	45
4.5	Pembagian Dataset.....	47

4.6	Pembangunan Model Klasifikasi .....	48
4.6.1.	Decision Tree .....	48
4.6.2.	Adaptive Boosting .....	49
4.6.3.	Gradient Boosting .....	50
4.7	Training.....	51
4.8	Validasi .....	52
4.9	Testing.....	53
4.10	Evaluasi Model .....	54
4.11	Hasil Validasi Model .....	55
4.12	Hasil Testing Model.....	59
4.13	<i>Confusion Matrix</i> .....	64
4.14	Hasil Perbandingan .....	67
4.15	Analisis Pengaruh <i>Learning Rate</i> .....	69
<b>BAB V PENUTUP.....</b>		<b>71</b>
5.1	Kesimpulan .....	71
5.2	Saran .....	72
<b>REFERENSI.....</b>		<b>73</b>

## DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian .....	9
Tabel 2. 2 Confusion Matrix .....	24
Tabel 3. 1 Rencana Perbandingan.....	40
Tabel 4. 1 Report Validasi Decision Tree.....	55
Tabel 4. 2 Report Validasi AdaBoost Konfigurasi Pertama .....	56
Tabel 4. 3 Report Validasi Adaboost Konfigurasi Kedua .....	57
Tabel 4. 4 Report Validasi Gradient Boosting Konfigurasi Pertama.....	58
Tabel 4. 5 Report Validasi Gradient Boosting Konfigurasi Kedua .....	59
Tabel 4. 6 Report Testing Decision Tree .....	60
Tabel 4. 7 Report Testing AdaBoost Konfigurasi Pertama .....	60
Tabel 4. 8 Report Testing AdaBoost Konfigurasi Kedua .....	61
Tabel 4. 9 Report Testing Gradient Boosting Konfigurasi Pertama.....	62
Tabel 4. 10 Report Testing Gradient Boosting Konfigurasi Kedua.....	63
Tabel 4. 11 Hasil Perbandingan .....	68
Tabel 4. 12 Perbandingan Performa Berdasarkan Learning rate .....	69

## DAFTAR GAMBAR

Gambar 2. 1 Struktur Decision Tree .....	17
Gambar 2. 2 Alur Kerja Adaptive Boosting .....	19
Gambar 3. 1 Dataset Url <i>Phishing</i> .....	26
Gambar 3. 2 Alur Penelitian .....	28
Gambar 3. 3 Tahapan Pre-Processing .....	30
Gambar 3. 4 Pembagian Data .....	31
Gambar 3. 5 Alur AdaBoost .....	33
Gambar 3. 6 Alur Gradient Boosting .....	35
Gambar 4. 1 Library .....	44
Gambar 4. 2 Proses Data Cleaning .....	46
Gambar 4. 3 Proses Data Transformation .....	46
Gambar 4. 4 Proses Feature Selection .....	47
Gambar 4. 5 Pembagian Data .....	48
Gambar 4. 6 Model Decision Tree .....	49
Gambar 4. 7 Model Adaboost Parameter Learning rate 0.1 .....	50
Gambar 4. 8 Model Adaboost Parameter Learning rate 1 .....	50
Gambar 4. 9 Model Gradient Boosting Parameter Learning rate 0.1 .....	51
Gambar 4. 10 Model Gradient Boosting Parameter Learning rate 1 .....	51
Gambar 4. 11 Training Decision Tree .....	51
Gambar 4. 12 Training Adaboost .....	52
Gambar 4. 13 Training Gradient Boosting .....	52
Gambar 4. 14 Tahap Validasi .....	53
Gambar 4. 15 Tahap Testing .....	54
Gambar 4. 16 Confusion Matrix Decision Tree .....	64
Gambar 4. 17 Confusion Matrix Adaboost .....	65
Gambar 4. 18 Confusion Matrix Gradient Boosting .....	66

## DAFTAR LAMBANG DAN SINGKATAN

$S$	Himpunan Kasus
$A$	Atribut
$N$	Jumlah partisi atribut
$P_i$	Proporsi dari $S_i$ terhadap $S$
$n$	Jumlah data latih
$M$	Jumlah iterasi
$w_i^{(m)}$	Bobot data ke- $i$ pada iterasi ke- $m$
$G_m(x)$	Model klasifikasi lemah ke- $m$
$\varepsilon_m$	tingkat kesalahan klasifikasi
$\alpha_m$	Bobot model
$Y$	Hasil klasifikasi akhir
$L(y_i, c)$	Fungsi kerugian
$rim$	Nilai residual atau sisa kesalahan pada iterasi ke- $m$
URL	Uniform Resource Locator
AdaBoost	Adaptive Boosting
TF-IDF	Term Frequency-Inverse Document Frequency
XGBoost	Extreme Gradient Boosting
CFS	Correlation-based Feature Selection
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
CSV	Comma Separated Values
RAM	Random Access Memory
SSD	Solid State Drive
VGA	Video Graphics Array

## DAFTAR ISTILAH

Phishing	Teknik penipuan yang memanfaatkan URL palsu untuk mencuri informasi sensitif pengguna dengan menyamar sebagai situs resmi.
Legitimate	Situs atau entitas resmi yang sah dan dapat dipercaya, yang tidak mengandung unsur penipuan serta beroperasi sesuai dengan aturan dan ketentuan yang berlaku.
Machine Learning	Bagian dari kecerdasan buatan yang memungkinkan sistem komputer belajar dari data dan mengidentifikasi pola tanpa diprogram secara eksplisit.
Supervised Learning	Pendekatan pembelajaran mesin yang menggunakan data berlabel untuk mengenali pola dan memprediksi data baru.
Decision Tree	Algoritma pembelajaran yang menggunakan struktur pohon serupa flowchart untuk tugas klasifikasi atau regresi.
Ensemble Learning	Teknik pembelajaran yang menggabungkan beberapa model (weak learners) untuk menciptakan satu model prediksi yang lebih kuat dan stabil.
Boosting	Metode ensemble yang membangun model secara berurutan, di mana model baru fokus memperbaiki kesalahan dari model sebelumnya.
Preprocessing	Tahap persiapan data untuk memastikan kualitas, konsistensi, dan kesesuaian format sebelum diproses oleh algoritma.
Overfitting	Kondisi di mana model terlalu menghafal detail data latih sehingga kinerjanya menurun saat menghadapi data baru.

Feature Selection	Proses pemilihan fitur-fitur yang paling relevan dan informatif dari dataset untuk meningkatkan efisiensi model.
Entropy	Ukuran ketidakpastian atau ketidakteraturan dalam suatu kumpulan data yang digunakan dalam perhitungan Decision Tree.
Information Gain	Ukuran efektivitas suatu atribut dalam mengklasifikasikan data.
Weak Learner	Algoritma pembelajaran sederhana yang digunakan sebagai komponen dasar dalam metode boosting.
Confusion Matrix	Tabel yang digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan hasil prediksi terhadap label sebenarnya.
Accuracy	Metrik yang menunjukkan persentase total prediksi benar dari keseluruhan data
Precision	Metrik yang mengukur tingkat ketepatan model dalam memprediksi kelas positif.
Recall	Metrik yang mengukur sensitivitas model dalam menemukan semua instans positif yang ada.
F1-Score	Rata-rata harmonik antara Precision dan Recall yang digunakan sebagai metrik keseimbangan performa.
Lexical Features	Karakteristik tekstual atau struktur karakter dari sebuah URL, seperti panjang karakter dan keberadaan simbol khusus

## INTISARI

Seiring dengan meningkatnya penggunaan internet, ancaman serangan siber berupa URL *phishing* semakin berkembang dan berpotensi mencuri informasi sensitif pengguna. Untuk mengatasi permasalahan tersebut, pendekatan *machine learning* dapat digunakan untuk mengklasifikasikan URL ke dalam kategori *phishing* dan *legitimate*. Penelitian ini bertujuan untuk meningkatkan performa algoritma Decision Tree melalui penerapan metode *ensemble learning* berbasis boosting, yaitu Adaptive Boosting dan Gradient Boosting.

Dataset yang digunakan terdiri dari dua kelas yaitu *phishing* dan *legitimate* yang telah melalui tahap preprocessing dan transformasi ke dalam fitur numerik agar dapat diproses oleh model klasifikasi. Tahapan penelitian meliputi pembagian data menjadi training, validasi, dan testing, proses pelatihan model, penyesuaian parameter seperti *learning rate*, serta evaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score.

Hasil penelitian menunjukkan bahwa model Decision Tree tanpa boosting memperoleh akurasi sebesar 0,94 dengan nilai presisi, recall, dan F1-score sebesar 0,94. Setelah diterapkan Adaptive Boosting, akurasi meningkat menjadi 0,97 dengan peningkatan yang konsisten pada seluruh metrik evaluasi. Model Gradient Boosting juga menghasilkan akurasi sebesar 0,97 dengan performa evaluasi yang seimbang. Hasil ini menunjukkan bahwa metode boosting efektif dalam meningkatkan akurasi dan kemampuan generalisasi Decision Tree dalam mendeteksi URL *phishing*.

**Kata kunci:** *URL phishing*, Decision Tree, Adaptive Boosting, Gradient Boosting, *Machine Learning*.

## **ABSTRACT**

*As internet usage increases, the threat of cyberattacks in the form of phishing URLs is growing, potentially stealing sensitive user information. To address this issue, a machine learning approach can be used to classify URLs into phishing and legitimate categories. This research aims to improve the performance of the Decision Tree algorithm by applying boosting-based ensemble learning methods, namely Adaptive Boosting and Gradient Boosting.*

*The dataset used consists of two classes phishing and legitimate, which have undergone preprocessing and transformation into numerical features for processing by the classification model. The research stages include dividing the data into training, validation, and testing, model training, parameter adjustments such as learning rate, and evaluation using accuracy, precision, recall, and F1-score metrics.*

*The results showed that the Decision Tree model without boosting achieved an accuracy of 0.94 with precision, recall, and F1-score values of 0.94. After applying Adaptive Boosting, the accuracy increased to 0.97 with consistent improvements across all evaluation metrics. The Gradient Boosting model also produced an accuracy of 0.97 with balanced evaluation performance. These results indicate that the boosting method is effective in improving the accuracy and generalization ability of Decision Trees in detecting phishing URLs.*

**Keyword:** *URL phishing, Decision Tree, Adaptive Boosting, Gradient Boosting, Machine Learning.*