L A P O R A N S K R I P S I

KOMPARASI ALGORITMA *NAIVE BAYES, SIMPLE LOGISTIC, RANDOM FOREST,* DAN *DECISION TREE J48* DALAM KLASIFIKASI SITUS WEB *PHISHING*

Nabilla Artamevia
NIM. 202151044

DOSEN PEMBIMBING

Alif Catur Murti, S.Kom, M.Kom

Rizkysari Meimaharani, S.Kom, M.Kom

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MURIA KUDUS
2025

L A P O R A N S K R I P S I

KOMPARASI ALGORITMA *NAIVE BAYES, SIMPLE LOGISTIC, RANDOM FOREST,* DAN *DECISION TREE J48* DALAM KLASIFIKASI SITUS WEB *PHISHING*

Nabilla Artamevia
NIM. 202151044

DOSEN PEMBIMBING

Alif Catur Murti, S.Kom, M.Kom

Rizkysari Meimaharani, S.Kom, M.Kom

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MURIA KUDUS
2025

HALAMAN PERSETUJUAN

KOMPARASI ALGORITMA *NAIVE BAYES, SIMPLE LOGISTIC, RANDOM FOREST,* DAN *DECISION TREE J48* DALAM KLASIFIKASI SITUS WEB *PHISHING*

NABILLA ARTAMEVIA

NIM. 202151044

Kudus, 19 Desember 2024

Menyetujui,

Pembimbing Utama,	Pembimbing Pendamping,
Alif Catur Murti, S.Kom., M.Kom	**Rizkysari Meimaharani, S.Kom., M.Kom**
NIDN. 0629077402	NIDN. 0620058501

Koordinator Skripsi,

Alvin Rainaldy Hakim, S.Kom., M.Kom

NIDN. 9990586218

HALAMAN PENGESAHAN

KOMPARASI ALGORITMA *NAIVE BAYES, SIMPLE LOGISTIC, RANDOM FOREST,* DAN *DECISION TREE J48* DALAM KLASIFIKASI SITUS WEB *PHISHING*

NABILLA ARTAMEVIA

NIM. 202151044

Kudus, 20 Januari 2025

Menyetujui,

Ketua Penguji,	Anggota Penguji I,	Anggota Penguji II,
Dr. Ahmad Abdul Chamid., M.Kom.	**Endang Supriyati., M.Kom**	**Alif Catur Murti., M.Kom**
NIDN. 0616109101	NIDN. 0629077402	NIDN. 0610129001

Mengetahui

Dekan Fakultas Teknik	Ketua Program Studi Teknik Informatika
Dr. Eko Darmanto, S.Kom., M.Cs	**Ir. Muhammad Imam Ghozali., M.Kom**
NIY. 0610701000001171	NIY. 0610701000001289

PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini:

Nama	:	Nabilla Artamevia
NIM	:	202151044
Tempat & Tanggal Lahir	:	Kudus, 19 Desember 2003
Judul Skripsi	:	Komparasi Algoritma Naive Bayes, Simple Logistic, Random Forest, Dan Decision Tree J48 Dalam Klasifikasi Situs Web Phishing

Dengan ini, saya menyatakan dengan sungguh-sungguh bahwa seluruh isi skripsi ini, mulai dari perumusan masalah, analisis data, hingga penulisan laporan akhir, merupakan hasil karya intelektual saya sendiri. Saya telah melakukan penelitian secara mandiri dan bertanggung jawab atas setiap informasi yang tertuang dalam skripsi ini. Segala ide, pendapat, atau temuan yang berasal dari sumber lain telah saya kutip dan rujuk sesuai dengan kaidah penulisan ilmiah yang berlaku.

Saya menyatakan dengan sungguh-sungguh bahwa seluruh isi skripsi ini merupakan hasil karya orisinal saya. Saya telah menjunjung tinggi prinsip-prinsip integritas akademik dalam proses penulisannya. Segala sumber yang digunakan telah dikutip dan dicantumkan dengan benar. Apabila di kemudian hari ditemukan adanya penyimpangan atau ketidaksesuaian dalam pernyataan ini, saya bersedia menerima segala bentuk sanksi akademik yang berlaku di Universitas Muria Kudus

Demikian pernyataan ini saya buat dalam keadaan sadar tanpa paksaan dari pihak manapun.

Kudus, 19 Desember 2024

Yang memberi pernyataan,

Nabilla Artamevia
NIM. 202151044

KATA PENGANTAR

Syukur Alhamdulillah, segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "Komparasi Algoritma Naive Bayes, Simple Logistic, Random Forest, dan Decision Tree J48 dalam Klasifikasi Situs Web Phishing". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik (S1) pada Program Studi Teknik Informatika.

Proses penyusunan skripsi ini tidak terlepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Ir. Darsono, M.Si., selaku Rektor Universitas Muria Kudus, yang telah memberikan fasilitas dalam proses studi hingga penyelesaian skripsi ini.
2. Bapak Ir. Muhammad Imam Ghozali, M.Kom, selaku Ketua Program Studi Teknik Informatika, yang telah memberikan arahan dan dukungan.
3. Bapak Alif Catur Murti, S.Kom., M.Kom, selaku pembimbing utama, yang dengan penuh kesabaran dan kebijaksanaan memberikan bimbingan, saran, dan kritik dalam penyelesaian skripsi ini.
4. Ibu Rizkysari Mei Maharani, S.Kom., M.Kom, atas masukan dan dukungannya selama proses penelitian dan penulisan skripsi ini.
5. Ibu Indah Fajarwati yang selalu memberikan doa, dukungan, serta motivasi yang tiada henti kepada penulis.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik, saran, dan masukan yang membangun dari para pembaca sangat penulis harapkan demi penyempurnaan karya di masa mendatang.

Akhir kata, penulis berharap skripsi ini dapat memberikan kontribusi kecil bagi pengembangan ilmu pengetahuan, khususnya dalam bidang klasifikasi situs web phishing.

Kudus, 19 Desember 2024

Penulis

JUDUL SKRIPSI
KOMPARASI ALGORITMA *NAIVE BAYES, SIMPLE LOGISTIC, RANDOM FOREST,* DAN *DECISION TREE J48* DALAM KLASIFIKASI SITUS WEB *PHISHING*

Nama mahasiswa : Nabilla Artamevia

NIM : 202151044

Pembimbing :

1. Alif Catur Murti, S.Kom, M.Kom
2. Rizkysari Meimaharani, S.Kom, M.Kom

RINGKASAN

Phishing adalah salah satu ancaman keamanan siber terbesar yang bertujuan mencuri informasi pribadi pengguna melalui situs web palsu. Penelitian ini bertujuan untuk membandingkan kinerja empat algoritma machine learning, yaitu Naive Bayes, Simple Logistic, Random Forest, dan Decision Tree J48, dalam mendeteksi situs web *phishing*. Selain itu, penelitian ini juga mengidentifikasi fitur paling signifikan yang berkontribusi dalam klasifikasi situs *phishing*, dengan harapan dapat meningkatkan efektivitas deteksi *phishing*.

Metode penelitian melibatkan penggunaan dataset berisi 11.055 data situs web yang diolah dengan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) untuk mengatasi ketidakseimbangan kelas. Evaluasi algoritma dilakukan menggunakan 10-Fold Cross Validation dan metrik evaluasi seperti akurasi, presisi, recall, F1 Score, MCC, dan ROC AUC. Setiap algoritma diuji untuk menentukan algoritma terbaik dalam mendeteksi situs phishing berdasarkan performa metrik tersebut.

Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan performa terbaik, dengan akurasi 97,60%, presisi 97,14%, recall 98,08%, dan F1 Score 97,61%. Fitur *SSLfinal_State* ditemukan sebagai atribut paling signifikan dalam mendeteksi situs *phishing*. Penelitian ini diharapkan dapat menjadi referensi bagi pengembangan sistem deteksi *phishing* yang lebih andal dan akurat.

Kata kunci: Phishing, Naive Bayes, Simple Logistic, Random Forest, Decision Tree J48.

COMPARISON OF NAIVE BAYES, SIMPLE LOGISTIC, RANDOM FOREST, AND DECISION TREE J48 ALGORITHMS IN CLASSIFYING PHISHING WEBSITES

Student Name : Nabilla Artamevia

Student Identity Number : 202151044

Supervisor :

1. Alif Catur Murti, S.Kom, M.Kom
2. Rizkysari Meimaharani, S.Kom, M.Kom

ABSTRACT

Phishing is one of the most significant cybersecurity threats that aims to steal users' private information through fraudulent websites. This research aims to compare the performance of four machine learning algorithms, namely Naive Bayes, Simple Logistic, Random Forest, and Decision Tree J48, in detecting phishing websites. Additionally, this study identifies the most significant features contributing to phishing classification to enhance detection effectiveness.

The research methodology involved using a dataset containing 11,055 website entries processed with the SMOTE (Synthetic Minority Over-sampling Technique) method to address class imbalance. Algorithm evaluation was conducted using 10-Fold Cross Validation and performance metrics such as accuracy, precision, recall, F1 Score, MCC, and ROC AUC. Each algorithm was tested to determine the best algorithm for phishing detection based on these performance metrics.

The results indicate that the Random Forest algorithm achieves the best performance, with 97.60% accuracy, 97.14% precision, 98.08% recall, and 97.61% F1 Score. The SSLfinal_State feature was identified as the most significant attribute in detecting phishing websites. This research is expected to serve as a reference for developing more reliable and accurate phishing detection systems.

***Keywords**: Phishing, Naive Bayes, Simple Logistic, Random Forest, Decision Tree J48.*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR SIMBOL

Simbol	Keterangan
P	Probabilitas atau kemungkinan yang terjadi.
$\sum$	Operator penjumlahan, digunakan untuk menghitung total nilai dalam suatu kumpulan data.
$\in$	Digunakan untuk menunjukkan bahwa suatu elemen termasuk dalam suatu himpunan.
argmax	Operator yang memaksimalkan fungsi tertentu.
$\prod$	Operasi perkalian berulang dari sekumpulan elemen dalam rentang tertentu.
|	Simbol untuk probabilitas bersyarat, berarti "dengan asumsi".
=	Simbol kesetaraan, menunjukkan bahwa kedua sisi memiliki nilai yang sama.
β_0	Intersep atau konstanta dalam model regresi logistik.
β_1	Koefisien regresi yang menunjukkan pengaruh variabel X terhadap Y.
-	Operasi pengurangan.
+	Operasi penjumlahan.
θ	Parameter yang digunakan untuk mengestimasi nilai dalam model machine learning.
$\neq$	Tidak sama dengan, digunakan untuk menyatakan ketidaksamaan antara dua nilai.
Θ	Parameter dalam model statistik atau machine learning.
$\leq$	Lebih kecil atau sama dengan, digunakan untuk membandingkan dua nilai.
ρ	Koefisien korelasi yang mengukur hubungan antara dua variabel.
log	Logaritma, operasi matematika untuk menemukan eksponen basis tertentu yang menghasilkan nilai tertentu.

DAFTAR LAMPIRAN

DAFTAR ISTILAH DAN SINGKATAN

SMOTE	:	Synthetic Minority Over-sampling Technique. Teknik untuk menangani ketidakseimbangan kelas dalam dataset dengan membuat data sintetis pada kelas minoritas.
ROC AUC	:	Receiver Operating Characteristic Area Under the Curve. Metrik evaluasi untuk mengukur kemampuan model membedakan antara kelas positif dan negatif.
MCC	:	Matthews Correlation Coefficient. Koefisien yang mengukur hubungan antara prediksi model dan kelas sebenarnya.
K-Fold Cross Validation	:	Teknik validasi model dengan membagi data ke dalam beberapa subset untuk pelatihan dan pengujian bergantian.
Phishing	:	Aktivitas kriminal yang bertujuan mencuri informasi pribadi pengguna melalui situs web palsu.
SSL	:	Secure Sockets Layer. Teknologi keamanan untuk mengenkripsi data yang dikirimkan antara server dan browser.
DNS	:	Domain Name System. Sistem yang menghubungkan nama domain dengan alamat IP.
URL	:	Uniform Resource Locator. Alamat yang digunakan untuk mengakses sumber daya di internet.
Redirect	:	Mekanisme untuk mengarahkan pengguna dari satu URL ke URL lainnya.
HTTPS	:	Hypertext Transfer Protocol Secure. Protokol keamanan untuk transfer data melalui internet.
F1 Score	:	Rata-rata harmonis antara presisi dan recall, digunakan sebagai metrik kinerja model klasifikasi.
HTML	:	Hypertext Markup Language. Bahasa markup untuk membuat struktur halaman web.
IP	:	Internet Protocol. Protokol komunikasi untuk mengidentifikasi perangkat di jaringan.
WHOIS	:	Layanan pencarian yang memberikan informasi mengenai kepemilikan nama domain.
PageRank	:	PageRank
Anchor Tag	:	Elemen HTML (<a>) yang digunakan untuk membuat hyperlink pada halaman web.

Iframe	:	Elemen HTML untuk menyisipkan dokumen atau konten dari halaman lain ke dalam halaman web.
Favicon	:	Ikon grafis kecil yang muncul di tab browser, biasanya merepresentasikan identitas situs web.
SMB	:	Server Message Block. Protokol jaringan yang memungkinkan berbagi file, printer, dan sumber daya lainnya.
PCA	:	Principal Component Analysis. Teknik pengurangan dimensi dalam dataset untuk meningkatkan efisiensi algoritma.
JavaScript	:	Bahasa pemrograman yang digunakan untuk membuat halaman web interaktif.