

**MACHINE LEARNING MODEL HYBRID UNTUK PROSES
DETEKSI SERANGAN WEB PHISING**

SKRIPSI

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



disusun oleh

ALYUDI RISAL

19.83.0406

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023**

**MACHINE LEARNING MODEL HYBRID UNTUK PROSES
DETEKSI SERANGAN WEB PHISING**

SKRIPSI

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



disusun oleh
ALYUDI RISAL
19.83.0406

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023

HALAMAN PERSETUJUAN
SKRIPSI
MACHINE LEARNING MODEL HYBRID UNTUK PROSES DETEKSI
SERANGAN WEB PHISING

yang disusun dan diajukan oleh

ALIYUDI RISAL

19.83.0406

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 15 Agustus 2023

Dosen Pembimbing,



Anggit Ferdita Nugraha, S.T., M.Eng.
NIK. 190302480

HALAMAN PENGESAHAN
SKRIPSI
MACHINE LEARNING MODEL HYBRID UNTUK PROSES DETEKSI SERANGAN WEB
PHISING

yang disusun dan diajukan oleh

ALYUDI RISAL

19.83.0406

Telah dipertahankan di depan Dewan Penguji
pada tanggal 15 Agustus 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Kusnawj, S.Kom, M.Eng.
NIK. 190302112

Jeki Kuswanto, M.Kom
NIK. 190302456

Anggit Ferdita Nugraha, S.T., M.Eng
NIK. 190302480

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 15 Agustus 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 1903020

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : ALIYUDI RISAL
NIM : 19.83.0406

Menyatakan bahwa Skripsi dengan judul berikut:

**MACHINE LEARNING MODEL HYBRID UNTUK PROSES DETEKSI
SERANGAN WEB PHISING**

Dosen Pembimbing : Anggit Ferdita Nugraha, S.T., 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 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, 15 Agustus 2023

Yang Menyatakan,



ALIYUDI RISAL

HALAMAN PERSEMBAHAN

Alhamdulillah hirobbil alamin, puji syukur atas nikmat yang telah diberikan Allah SWT sehingga penulis dapat menyelesaikan skripsi ini. Dengan bangga saya mempersembahkan hasil Skripsi ini untuk orang tua dan saudara serta teman-teman saya yang senantiasa memberi dukungan tiada henti untuk saya serta kasih sayangnya, sehingga penulis dapat menggapai tujuan hidup dan menjalani hidup dengan penuh anugerah.



KATA PENGANTAR

Puji syukur saya panjatkan kepada Allah SWT Tuhan Yang Maha Esa, dengan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Skripsi dengan judul: "MACHINE LEARNING MODEL HYBRID UNTUK PROSES DETEKSI SERANGAN WEB PHISING". Shalawat serta salam tidak lupa senantiasa dilimpahkan kepada Nabi Muhammad Shallallahu 'Alaihi Wasallam, yang telah membawa kita dari zaman jahiliyah menuju ke zaman terang benderang. Sebagai wujud rasa syukur penulis atas selesainya penulisan Skripsi ini maka penulis ingin mengucapkan terimakasih kepada:

1. Bapak Anggit Ferdita Nugraha, ST., M.Eng, selaku pembimbing yang telah memberikan arahan serta bimbingan sehingga penulis mampu menyelesaikan skripsi ini.
2. Orang tua saya yang senantiasa memberikan dukungan yang tiada hentinya kepada saya.

Yogyakarta, 8 Agustus 2023



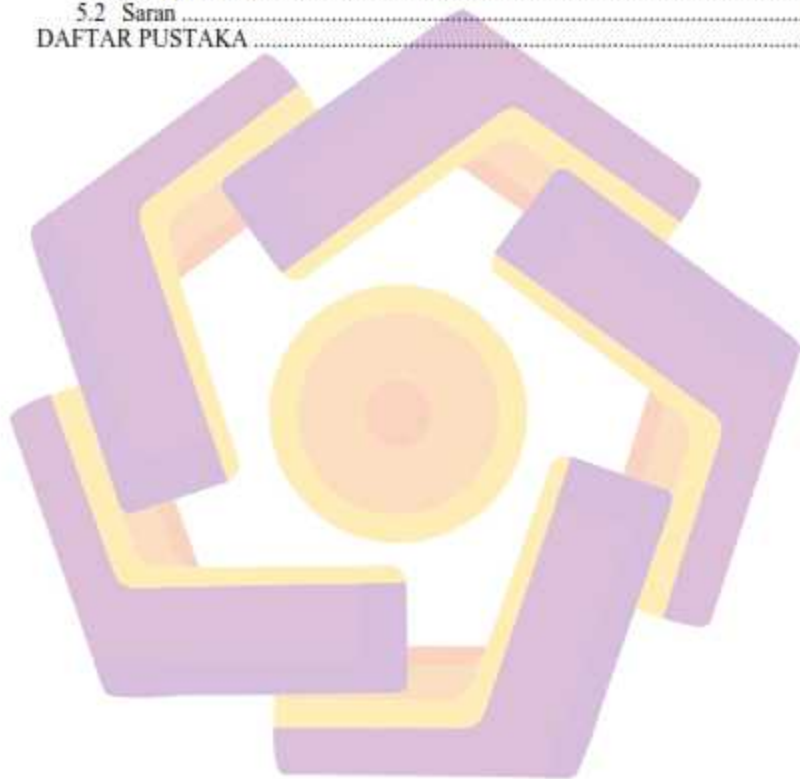
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN SKRIPSI.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
INTISARI	xiii
ABSTRACT	xiv
BAB I PENDAHULUAN	15
1.1 Latar Belakang	15
1.2 Rumusan Masalah	16
1.3 Batasan Masalah	16
1.4 Tujuan Penelitian	17
1.5 Manfaat Penelitian	17
BAB II TINJAUAN PUSTAKA	18
2.1 Studi Literatur	18
2.2 Dasar Teori	22
2.2.1 Web Phishing	22
2.2.2 Artificial Intelligence	22
2.2.3 Machine Learning	23
2.2.3.1 Supervised Learning	24
2.2.3.2 Unsupervised Learning	24
2.2.3.3 Semi Supervised Learning	25
2.2.3.4 Reinforcement Learning	25

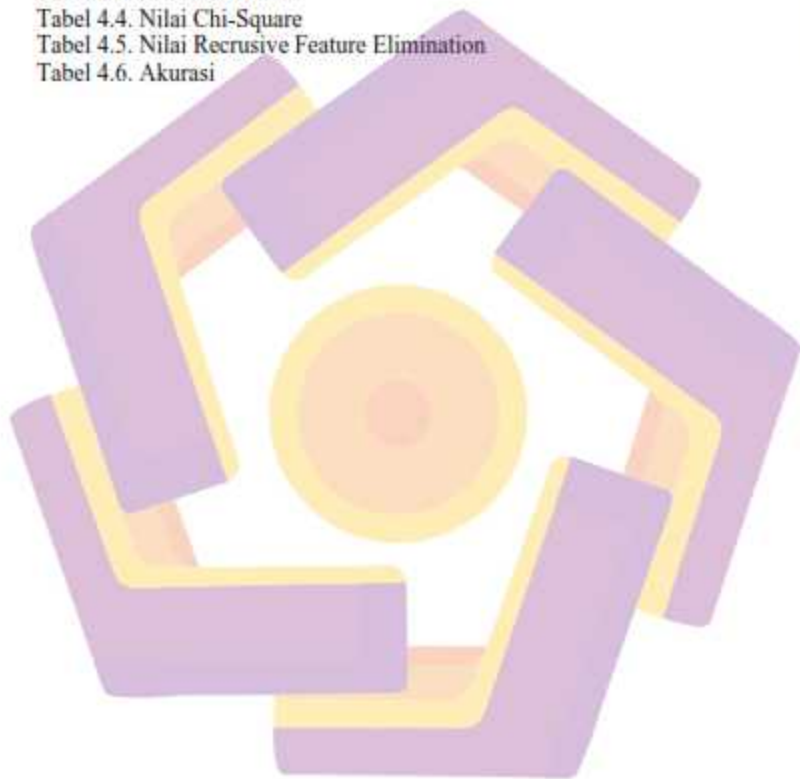
2.2.4	Klasifikasi	25
2.2.5	Decision Tree	26
2.2.6	Fitur Seleksi	26
2.2.7	Recursive Feature Elimination	26
2.2.8	Gain Ratio	27
2.2.9	Pearson Correlation	28
2.2.10	Chi-Square	29
2.2.11	Information Gain	30
2.2.12	Ensembled Model	31
2.2.13	Meta Algoritma	32
2.2.13.1	Bagging	32
2.2.13.2	Boosting	32
2.2.13.3	Stacking	33
2.2.14	Evaluasi	33
2.2.14.1	Akurasi	34
2.2.14.2	Presisi	34
2.2.14.3	Recall	34
2.2.14.4	F1 Score	34
2.2.15	Ruang Lingkup Ilmu	35
2.2.16	Skenario	36
BAB III METODE PENELITIAN		40
3.1	Alat dan Bahan	40
3.2	Langkah Penelitian	40
3.2.1	Data Acquisition	41
3.2.2	Feature Engineering	48
3.2.3	Klasifikasi	48

3.2.4 Evaluasi	49
BAB IV HASIL DAN PEMBAHASAN	51
5.2 Implementasi	51
5.2 Pengujian	58
BAB V PENUTUP	83
5.1 Kesimpulan	83
5.2 Saran	83
DAFTAR PUSTAKA	85



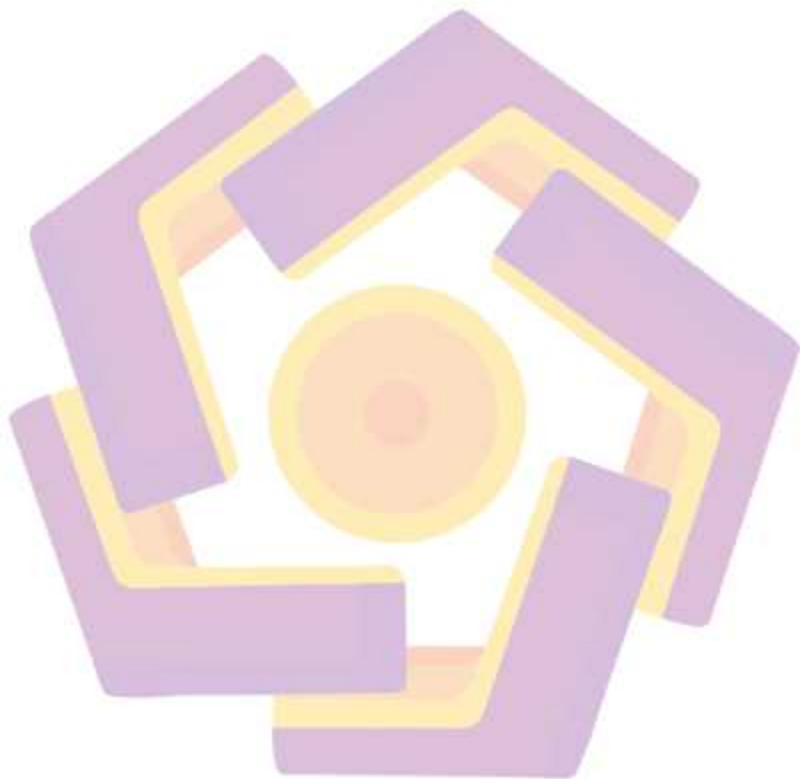
DAFTAR TABEL

Tabel 2.1. Referensi Penelitian	17
Tabel 3.1. Fitur Web Phishing	37
Tabel 4.1. Nilai Korelasi Pearson Corelation	48
Tabel 4.2. Nilai Information Gain	49
Tabel 4.3. Nilai Gain Ratio	50
Tabel 4.4. Nilai Chi-Square	50
Tabel 4.5. Nilai Recursive Feature Elimination	52
Tabel 4.6. Akurasi	76



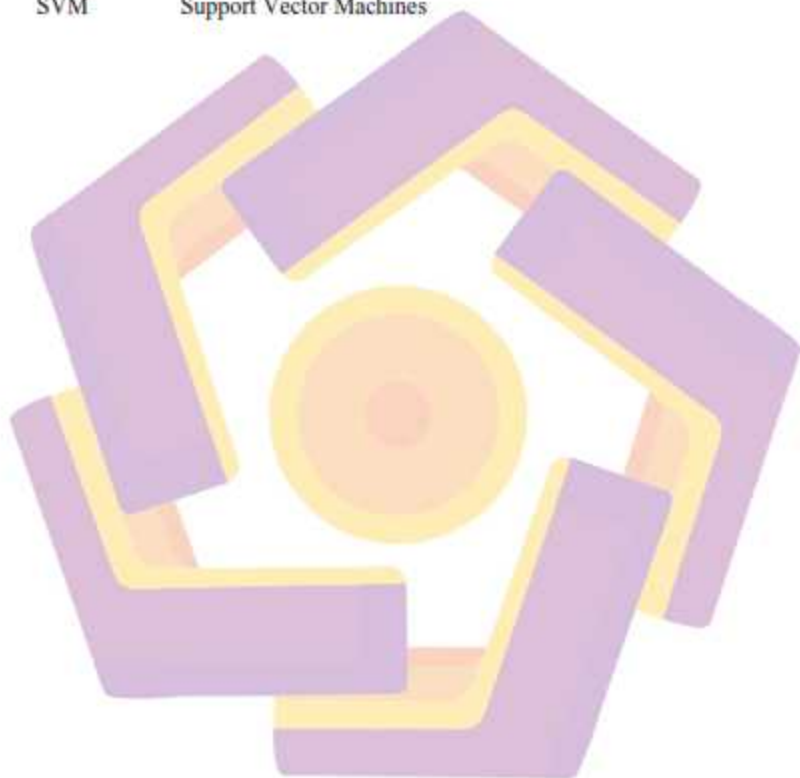
DAFTAR GAMBAR

Gambar 2.1. Artificial Intelligence	22
Gambar 2.2. Machine Learning	23
Gambar 3.1. Alur Penelitian	37



DAFTAR LAMBANG DAN SINGKATAN

Ω	Tahanan Listrik
μ	Konstanta gesekan
ANFIS	Adaptive Network Fuzzy Inference System
SVM	Support Vector Machines



INTISARI

Web phishing merupakan salah satu bentuk kejahatan siber yang mengancam aktivitas online pengguna situs web. Dalam serangan ini, para pelaku membuat halaman web tiruan yang sangat menyerupai situs resmi untuk memancing korban agar memberikan informasi penting. Dalam beberapa tahun terakhir, serangan phishing web telah mengalami peningkatan yang signifikan, memicu kebutuhan untuk mengembangkan sistem deteksi guna mengurangi jumlah korban dan kerugian finansial akibat serangan ini. Sistem deteksi phishing web telah menjadi fokus penelitian dan pengembangan dengan penerapan metode pembelajaran mesin. Namun, sebagian besar pendekatan ini hanya terfokus pada tahap klasifikasi, tanpa mempertimbangkan pentingnya seleksi fitur yang tepat. Dalam kajian ini, eksperimen dilakukan dengan menerapkan langkah fitur seleksi berdasarkan lima algoritma populer: pearson correlation, information gain, gain ratio, chi-square, dan recursive feature elimination. Langkah ini diambil sebelum melakukan pembelajaran mesin menggunakan algoritma yang umum digunakan, seperti Decision Tree, dan dilakukan juga penambahan meta-algoritma untuk memungkinkan peningkatan kinerja klasifikasi. Dalam konteks ini, hasil akurasi dari seleksi fitur recursive feature elimination mencapai 97,9%, sedangkan penggunaan meta-algoritma seperti bagging dengan seleksi fitur recursive feature elimination memperoleh akurasi 94,6%. Selain itu, metode boosting mencapai akurasi 93,2%, dan metode stacking mencapai akurasi 93,0%. Dari hasil kinerja ini, terlihat bahwa penambahan meta-algoritma pada seleksi fitur recursive feature elimination tidak menghasilkan peningkatan signifikan dalam akurasi.

Kata kunci: Web phishing, fitur seleksi, machine learning, decision tree, meta-algoritma

ABSTRACT

Web phishing is one form of cybercrime that threatens the online activities of website users. In this type of attack, perpetrators create counterfeit web pages that closely resemble official sites to lure victims into providing crucial information. In recent years, web phishing attacks have experienced a significant increase, prompting the need to develop detection systems to reduce the number of victims and financial losses resulting from these attacks. Web phishing detection systems have been the focus of research and development, incorporating machine learning methods. However, most of these approaches have primarily concentrated on the classification stage without considering the importance of proper feature selection. In this study, experiments were conducted by implementing feature selection steps based on five popular algorithms: Pearson correlation, information gain, gain ratio, chi-square, and recursive feature elimination. These steps were taken prior to machine learning using commonly used algorithms, such as Decision Trees, and the addition of meta-algorithms was also carried out to enhance classification performance. In this context, the accuracy results from the recursive feature elimination feature selection reached 97.9%, while the use of meta-algorithms like bagging with recursive feature elimination feature selection achieved an accuracy of 94.6%. Additionally, the boosting method achieved an accuracy of 93.2%, and the stacking method reached an accuracy of 93.0%. From these performance outcomes, it can be observed that the addition of meta-algorithms to the recursive feature elimination feature selection did not yield a significant improvement in accuracy.

Keyword: Web phishing, feature selection, machine learning, decision tree, meta-algorithms.