

**IDENTIFIKASI WEBSITE PHISHING MENGGUNAKAN
ALGORITMA C4.5 DAN SUPPORT VECTOR MACHINE**

SKRIPSI

Diajukan untuk memenuhi salah satu persyaratan memperoleh gelar Sarjana
Komputer pada Program Studi Informatika



Disusun Oleh:

RANDI ARDHANI RAHMAWAN

19.11.2935

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

**IDENTIFIKASI WEBSITE PHISHING MENGGUNAKAN
ALGORITMA C4.5 DAN SUPPORT VECTOR MACHINE**

SKRIPSI

Diajukan untuk memenuhi salah satu persyaratan memperoleh gelar Sarjana
Komputer pada Program Studi Informatika



Disusun Oleh:

RANDI ARDHANI RAHMAWAN

19.11.2935

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

IDENTIFIKASI WEBSITE PHISHING MENGGUNAKAN ALGORITMA C4.5
DAN SUPPORT VECTOR MACHINE

Yang disusun dan diajukan oleh

RANDI ARDHANI RAHMAWAN

19.11.2938

Telah disetujui oleh Dosen Pembimbing Skripsi
Pada tanggal 25 Mei 2023

Dosen Pembimbing,



FERIAN FAUZI ABDULLOH, M.Kom
NIK. 196302276

HALAMAN PENGESAHAN
SKRIPSI
IDENTIFIKASI WEBSITE PHISHING MENGGUNAKAN ALGORITMA C4.5
DAN SUPPORT VECTOR MACHINE

Yang dibuat dan diajukan oleh

RANDI ARDHANI RAHMAWAN
19.11.2935

Telah dipertalumkan di depan Dewan Penguji
pada tanggal 23 Mei 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anna Bailla, M.Kom
NIK. 190302190

Jeki Kuswanto, M.Kom
NIK. 190302456

Ferhan Fauzi Abdullah, M.Kom
NIK. 190302276

Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Mei 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fala, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Randi Ardhani Rahmawan**

NIM : **19.11.2935**

Menyatakan bahwa Skripsi dengan judul berikut:

IDENTIFIKASI WEBSITE PHISHING MENGGUNAKAN ALGORITMA C4.5 DAN SUPPORT VECTOR MACHINE

Dosen Pembimbing : **Feriz Fauzi Abdullah, M.Kom**

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 Mei 2023

Yang Menyatakan,



Randi Ardhani Rahmawan

HALAMAN PERSEMBAHAN

Puji dan Syukur atas kehadiran Allah SWT, yang telah melimpahkan rahmat, hidayah dan karunia-Nya, atas kemudahan, petunjuk dan kekuatan yang diberikan-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul “Identifikasi Website Phishing Menggunakan Algoritma C4.5 Dan Support Vector Machine”, dan diajukan sebagai salah satu syarat untuk mendapat gelar Sarjana pada prodi Informatika, Universitas AMIKOM Yogyakarta. Karya tulis ini penulis persembahkan kepada :

1. Pertama-tama saya ucapkan terima kasih yang sebesar-besarnya kepada ayah saya Rohman Warsito, yang selalu mendukung dan menyemangati saya serta memberikan kebutuhan finansial saya selama menempuh pendidikan hingga menjadi seorang sarjana.
2. Kedua saya ucapkan terima kasih yang sebesar-besarnya kepada ibu saya Dinna Mariyanna, yang selalu mendukung dan menyemangati saya dalam menjalani pendidikan di Yogyakarta. Terima kasih telah mendoakan saya dan selalu mengingatkan saya untuk tidak meninggalkan ibadah lima waktu.
3. Ketiga saya ucapkan terima kasih kepada adik saya Meisya Azzahra Andini Rahmawan yang selalu menyemangati dan mendukung saya selama menjalani pendidikan di Yogyakarta.
4. Terakhir saya ucapkan terima kasih kepada semua pihak yang telah mendukung dan memberikan semangat selama proses penulisan skripsi ini dan juga selama menjalani pendidikan ini.

KATA PENGANTAR

Puji dan Syukur atas kehadiran Allah SWT yang telah melimpahkan rahmat, hidayah dan karunia-Nya, atas kemudahan, petunjuk dan kekuatan yang diberikan-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul "Identifikasi Website Phishing Menggunakan Algoritma C4.5 Dan Support Vector Machine", dan diajukan sebagai salah satu syarat untuk mendapat gelar Sarjana pada jurusan Informatika, Universitas Amikom Yogyakarta. Dalam penyusunannya, penulis memperoleh banyak bantuan dari berbagai pihak, oleh karena itu penulis mengucapkan terimakasih yang sebesar-besarnya kepada :

1. Bapak Ferian Fauzi Abdulloh, M. Kom sebagai dosen pembimbing yang telah meluangkan waktunya untuk memberikan saran dan juga membimbing penulis untuk menyelesaikan skripsi ini.
2. Seluruh dosen Prodi Informatika Universitas AMIKOM Yogyakarta yang telah memberikan ilmu kepada penulis.
3. Kedua orang tua dan keluarga besar yang telah mendoakan dan menyemangati dalam proses pengerjaan skripsi ini.
4. Kepada pemilik NIM 1951020194 yang telah mendukung dan memberikan semangat untuk menyelesaikan skripsi ini.
5. Teman-teman yang selalu memberikan semangat dalam proses pengerjaan skripsi ini.
6. Semua pihak yang telah membantu penulis dalam mengerjakan skripsi ini dari awal hingga akhir.

Tidak ada kata selain terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu saya sampai titik ini. Penulis menyadari bahwa penulisan skripsi ini masih memiliki banyak kekurangan. Untuk semua kritik dan saran yang bersifat membangun sangat penulis harapkan dan akan diterima dengan lapang dada.

Yogyakarta, 25 Mei 2023



Randi Ardhani Rahmawan

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
ABSTRACT	xvi
BAB I PENDAHULUAN	1
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
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	15
2.2.1 Data Mining	15
2.2.2 Tahap-Tahap Data Mining	15

2.2.3	Seleksi Fitur	16
2.2.4	Information Gain	16
2.2.5	C4.5	17
2.2.6	Support Vector Machine.....	19
2.2.7	K-Fold Cross Validation.....	22
2.2.8	Confusion Matrix.....	22
BAB III METODE PENELITIAN		24
3.1	Objek Penelitian.....	24
3.2	Alur Penelitian	24
3.2.1	Perumusan Masalah.....	25
3.2.2	Studi Literatur.....	25
3.2.3	Dataset.....	25
3.2.4	Preprocessing.....	25
3.2.5	C4.5	25
3.2.6	Support Vector Machine.....	27
3.2.7	Confusion Matrix.....	28
3.2.8	Kesimpulan.....	28
3.3	Alat dan Bahan.....	28
3.3.1	Data Penelitian.....	28
3.3.2	Alat Penelitian	29
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Dataset.....	30
4.2	Implementasi <i>Preprocessing</i>	31
4.2.1	Data <i>Cleaning</i>	31
4.2.2	Data <i>Reduction</i>	31
4.3	Uji Akurasi Algoritma C4.5.....	34
4.4	Uji Akurasi Algoritma Support <i>Vector Machine</i>	42

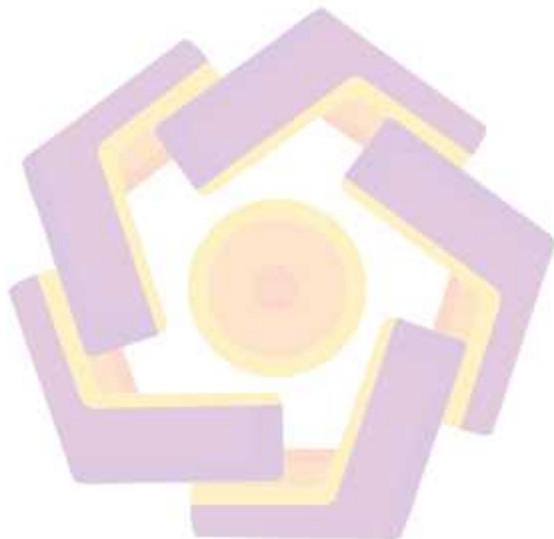
4.5 Perbandingan Akurasi Dengan Penelitian Terdahulu (Studi Literatur)	51
BAB V PENUTUP	53
5.1 Kesimpulan	53
5.2 Saran.....	53
DAFTAR PUSTAKA	54



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	9
Tabel 3.3 Ilustrasi Proses 3 Fold Cross Validation Algoritma C4.5.....	26
Tabel 3.4 Ilustrasi Proses 5 Fold Cross Validation Algoritma C4.5.....	26
Tabel 3.5 Ilustrasi Proses 10 Fold Cross Validation Algoritma C4.5.....	26
Tabel 3.6 Ilustrasi Proses 3 Fold Cross Validation Algoritma SVM.....	27
Tabel 3.7 Ilustrasi Proses 5 Fold Cross Validation Algoritma SVM.....	27
Tabel 3.8 Ilustrasi Proses 10 Fold Cross Validation Algoritma SVM.....	28
Tabel 3.1 Perangkat Lunak (Software)	29
Tabel 3.2 Perangkat Keras (Hardware).....	29
Tabel 4.1 Karakteristik Dari Dataset Website Phishing	30
Tabel 4.2 30 Fitur Awal Sebelum Dilakukannya Seleksi Atribut	30
Tabel 4.3 Seleksi Fitur Information Gain	32
Tabel 4.4 Total <i>Confusion Matrix</i> 3 Fold Cross Validation Pada JupyterLab	34
Tabel 4.5 Hasil Pengujian 3 Fold Cross Validation Pada JupyterLab.....	35
Tabel 4.6 Total <i>Confusion Matrix</i> 5 Fold Cross Validation Pada JupyterLab	35
Tabel 4.7 Hasil Pengujian 5 Fold Cross Validation JupyterLab.....	36
Tabel 4.8 Total <i>Confusion Matrix</i> 10 Fold Cross Validation Pada JupyterLab	36
Tabel 4.9 Hasil Pengujian 10 Fold Cross Validation JupyterLab.....	38
Tabel 4.10 Total <i>Confusion Matrix</i> 3 Fold Cross Validation Pada Weka	38
Tabel 4.11 Hasil Pengujian 3 Fold Cross Validation Weka	39
Tabel 4.12 Total <i>Confusion Matrix</i> 5 Fold Cross Validation Pada Weka	39
Tabel 4.13 Hasil Pengujian 5 Fold Cross Validation Weka	40
Tabel 4.14 Total <i>Confusion Matrix</i> 10 Fold Cross Validation Pada Weka	40
Tabel 4.15 Hasil Pengujian 10 Fold Cross Validation Weka	41
Tabel 4.16 Total <i>Confusion Matrix</i> 3 Fold Cross Validation Pada JupyterLab	42
Tabel 4.17 Hasil Pengujian 3 Fold Cross Validation JupyterLab.....	43
Tabel 4.18 Total <i>Confusion Matrix</i> 5 Fold Cross Validation Pada JupyterLab	44
Tabel 4.19 Hasil Pengujian 5 Fold Cross Validation JupyterLab.....	45
Tabel 4.20 Total <i>Confusion Matrix</i> 10 Fold Cross Validation Pada JupyterLab	45
Tabel 4.21 Hasil Pengujian 10 Fold Cross Validation JupyterLab.....	46
Tabel 4.22 Total <i>Confusion Matrix</i> 3 Fold Cross Validation Pada Weka	46

Tabel 4.23 Hasil Pengujian 3 Fold Cross Validation Weka	47
Tabel 4.24 Total <i>Confusion Matrix</i> 5 Fold Cross Validation Pada Weka	48
Tabel 4.25 Hasil Pengujian 5 Fold Cross Validation Weka	49
Tabel 4.26 Total <i>Confusion Matrix</i> 10 Fold Cross Validation Pada Weka	49
Tabel 4.27 Hasil Pengujian 10 Fold Cross Validation Weka	50
Table 4.28 Perbandingan Akurasi dengan penelitian Terdahulu	51



DAFTAR GAMBAR

Gambar 2.1 Tahapan-Tahapan Data Mining	15
Gambar 2.2 Ilustrasi SVM	19
Gambar 2.3 <i>Confusion Matrix</i>	23
Gambar 3.1 Alur Penelitian	24
Gambar 4.1 Perbandingan Akurasi Pada JupyterLab dan Weka C4.5	42
Gambar 4.2 Perbandingan Akurasi Pada JupyterLab dan Weka SVM	50



DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machine
APWG	Anti Phishing Working Group
BEC	Business Email Compromise
IDADX	Indonesia Anti-Phishing Data Exchange
IRS	International Revenue Service
ISP	Internet Service Provider
CNN	Convolutional Neural Networks
ANFIS	Adaptive Neuro-Fuzzy Inference System
FVV	Feature Validity Value
URL	Uniform Resource Locator
DT	Decision Tree
RF	Randon Forest
PSO	Particle Swarm Optimation
BPNN	Back-Propagation Neural Network
KNN	K-Nearest Neighbor
NB	Naïve Bayes
ML	Machine Learning
ID3	Iterative Dichotomiser 3
SRM	Structural Risk Minimization
TP	True Positive
FP	True Negative
TN	False Positive
FN	False Negative

DAFTAR ISTILAH

Preprocessing Pengolahan atau pemrosesan data



INTISARI

Indonesia ialah salah satu negara dengan pengguna internet terbesar di dunia. Menurut laporan *We Are Social*, terdapat 204.7 juta pengguna internet per Januari 2022. Pengguna internet tersebut kebanyakan digunakan untuk aktifitas sosial di dunia maya. Terbukti dengan jumlah penggunaan sosial media yang begitu besar jumlahnya, mencapai 191,4 juta pengguna pada Januari 2022. Dengan maraknya penggunaan internet di Indonesia, ancaman terhadap penggunanya semakin kian beragam. Salah satunya ialah phishing. Tujuan dari serangan tersebut ialah membuat penggunanya percaya bahwa mereka berinteraksi dengan situs resmi yang dimaksud. Umumnya informasi yang dicari phiser (pelaku phishing) yaitu informasi tentang kartu kredit korban atau pun password akun, dengan cara dialihkan ke web palsu atau dengan mengirimkan email ke korban, dimana korban akan diminta untuk memberikan informasi pribadinya. Tujuan dari penelitian ini adalah untuk mendapatkan tingkat akurasi dalam mengklasifikasi *website phishing* dan dapat menentukan jenis *website* termasuk dalam *website phishing* atau *website legitimate*. Penelitian ini akan menggunakan algoritma C4.5 dan Support Vector Machine kemudian menggunakan *Information Gain* untuk melakukan seleksi fitur. Hasil penelitian ini menunjukkan bahwa algoritma C4.5 memiliki akurasi sebesar 95.77%, presisi 94.74%, recall 92.32%, dan skor F1 93.52% sedangkan *Support Vector Machine* memiliki akurasi sebesar 94.33%, presisi 95.48%, recall 94.94%, dan skor F1 95.21%.

Kata kunci: Phising, Identifikasi, C4.5, Support Vector Machine



ABSTRACT

Indonesia is one of the countries with the largest internet users in the world. According to the We Are Social report, there are 204.7 million internet users as of January 2022. These internet users are mostly used for social activities in cyberspace. This is evidenced by the large number of social media users, reaching 191.4 million users in January 2022. With the widespread use of the internet in Indonesia, threats to its users are increasingly diverse. One of them is phishing. The aim of the attack is to make its users believe that they are interacting with the legitimate website in question. Generally, the information that phishers are looking for is information about the victim's credit card or account password, by being redirected to a fake website or by sending an email to the victim, where the victim will be asked to provide personal information. The purpose of this study is to obtain a level of accuracy in classifying websites and to determine the types of phishing websites included in phishing websites or legitimate websites. This research will use the C4.5 algorithm and Support Vector Machine then use Information Gain to perform feature selection. The results of this study indicate that the C4.5 algorithm has an accuracy of 95.77%, precision of 94.74%, recall of 92.32%, and f1 score of 93.52% while the Support Vector Machine has an accuracy of 94.33%, precision of 95.48%, recall of 94.94%, and f1 score of 95.21%

Keywords: *Phishing, Identification, C4.5, Support Vector Machine*