ANALISIS MALICIOUS URL MENGGUNAKAN TEKNIK DATA MINING

SKRIPSI

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

disusun oleh
DENIE WAHYU PRATAMA
18.83.0293

Kepada

FAKULTAS ILMU KOMPUTER
UNAIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023

ANALISIS MALICIOUS URL MENGGUNAKAN TEKNIK DATA MININIG

SKRIPSI

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

disusun oleh
DENIE WAHYU PRATAMA
18.83.0293

Kepada

FAKULTAS ILMU KOMPUTER
UNAIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023

HALAMAN PERSETUJUAN

SKRIPSI

ANALISIS MALICIOUS URL MENGGUNAKAN TEKNIK DATA MINING

yang disusun dan diajukan oleh

Denie Wahyu Pratama

18.83.0293

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 31 Maret 2023

Dosen Pembimbing,

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

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS MALICIOUS URL MENGGUNAKAN TEKNIK DATA MINING

yang disusun dan diajukan oleh

Denie Wahyu Pratama

18.83.0293

Telah dipertahankan di depan Dewan Penguji
pada tanggal 31 Maret 2023

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Yoga Pristyanto, S.Kom, M.Eng **NIK. 190302412**	
Nurhini, M.Kom **NIK. 190302066**	
Anggit Ferdita Nugraha, S.T, M.Eng **NIK. 190302480**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 31 Maret 2023

DEKAN FAKULTAS ILMU KOMPUTER

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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Denie Wahyu Pratama
NIM : 18.83.0293

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Malicious URL Menggunakan Teknik Data Mining

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, 31 Maret 2023

Yang Menyatakan,

Denie Wahyu Pratama

HALAMAN PERSEMBAHAN

Skripsi ini saya persembahkan dengan penuh rasa syukur kepada Allah SWT yang telah melimpahkan rahmat dan kesehatan serta petunjuk sehingga saya dapat menyelesaikan skripsi ini. Skripsi ini saya persembahkan kepada semua pihak yang telah membantu secara langsung maupun tidak langsung, yaitu saya persembahkan untuk :

1. Unaiversitas AMIKOM Yogyakarta.
2. Ayah dan Ibu, terima kasih atas doa, semangat, motivasi, pengorbanan, nasihat serta kasih sayang yang tidak pernah henti sampai saat ini.
3. Keluarga besar serta saudara.
4. Teman dan sahabat yang tidak bisa saya sebutkan satu persatu.
5. Semua pihak yang telah membantu skripsi ini.

KATA PENGANTAR

Dengan memanjatkan puji dan syukur kehadirat Allah SWT yang telah melimpahkan Rahmat dan Hidayahnya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul "*Analisis Malicious URL Menggunakan Teknik Data Mining*" dengan diberikan kelancaran dan kemudahan, walaupun disadari masih ada beberapa kekurangan yang tidak lepas karena keterbatasan penulis.

Skripsi ini di susun sebagai salah satu syarat kelulusan program sarjana mahasiswa/mahasiswi di Unaiversitas AMIKOM Yogyakata. Dalam peyusunan skipsi ini tak lepas dari hambatan dan kendala. Namun berkat tekad, usaha dan dorongan serta bantuan dari berbagai pihak yang akhirnya penulis dapat menyelesaikan skripsi ini. Oleh karena itu, dengan segala kerendahan hati dan tanpa mengurangi rasa hormat, pada kesempatan ini penulis menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Allah SWT Tuhan Yang Maha Esa, yang selalu memberikan Rahmat serta Hidayah-Nya, sehingga selalu diberikan kemudahan, kelancaran dalam mengerjakan tugas akhir.
2. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Unaiversitas AMIKOM Yogyakarta.
3. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Unaiversitas AMIKOM Yogyakarta.
4. Bapak Dony Ariyus, M.Kom. selaku Ketua Program Studi Teknik Komputer Unaiversitas Amikom Yogyakarta.
5. Bapak Anggit Ferdita Nugraha,S.T.,M.Eng selaku Dosen Pembimbing yang telah memotivasi, membimbing dan memberikan masukkan dalam penyelesaian skripsi ini.
6. Kedua Orang Tua, keluarga besar dan saudara yang selalu memberikan dukungan berupa moril dan materil, serta doa yang tak henti hentinya.
7. Rekan-rekan mahasiswa Unaiversitas AMIKOM Yogyakarta yang selalu memberikan dukungan, motivasi dan doa.

8. Semua pihak yang tidak dapat penulis sebutkan satu per satu tetapi telah banyak berjasa dalam memberikan bantuan moril maupun materil kepada penulis.

Penulis sadar bahwa skripsi ini jauh dari kata sempurna, masih banyak kekuranagan yang perlu di benahi. Maka dari itu penulis mengharapkan kritik dan saran yang membangun. Walupuan demikian penulis berharap skripsi ini dapat membantu dan memberikan manfaat bagi semua pihak baik penulis maupun pembaca.

Demikian apabila terdapat kesalahan dalam menulis nama dan kata, penulis sampaikan permintaan maaf yang sebesar-besarnya.

Yogyakarta, 31 Maret 2023

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMBANG DAN SINGKATAN

URL	Uniform Resource Locators
API	Application Programming Interface
CART	Classification And Regression Trees
HTTP	Hypertext Transfer Protocol
PHP	Hypertext Preprocessor
CSS	Cascading Style Sheet
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
WWW	World Wide Web
SVM	Support Vector Machine
KNN	K Nearst Neighbors
WSGI	Web Server Gateway Interface
SOAP	Simple Object Access Protocol
REST	Representational State Transfer

DAFTAR ISTILAH

Atribut	Karakteristik atau sifat yang melekat dari sebuah objek data.
Dataset	Kumpulan data-data mentah berupa tabel yang dapat dioleh lebih lanjut.
Decision Tree	Pohon keputusan
Protokol	Sejumlah aturan yang menentukan bagaimana dua atau lebih komputer saling berkomunikasi.
Web	Layanan internet yang digunakan untuk membagikan informasi secara global dan mudah diakses melalui jaringan internet.
Download	Pengambilan data atau informasi dari sebuah web site untuk ditampilkan atau disimpan di komputer pengguna.
Overfitting	Kondisi di mana model *machine learning* terlalu kompleks dan terlalu dekat dengan data latih. Artinya, model memiliki performa yang sangat baik pada data latih, tetapi buruk dalam prediksi data yang belum pernah dilihat sebelumnya.
String	Data berupa huruf
Null value	Nilai kosong
Black List	Daftar hitam

INTISARI

Pada pendeteksian website phishing metode konvensional yang paling banyak digunakan oleh vendor antivirus dan peneliti. Metode ini meliputi black list, white list dan signature yang masih mempunyai banyak kekurangan diantaranya tidak bisa mendeteksi web phishing baru maupun replika web yang mempunyai konten berbeda dan web yang berbahaya lainya. Beberapa peneliti memaksimalkan pendeteksiannya menggunakan algoritma data mining maupun machine learning dan berlomba lomba menghasilkan akurasi yang lebih baik. Oleh karena itu, pada penelitian ini diusulkan penggunaan machine learning dan data mining dengan algoritma decision tree untuk deteksi web phishing memanfaatkan data dari platform online API VirusTotal. Data yang diperoleh berupa status URL beserta vendor yang mendeteksinya yang berjumlah 819 URL. Penggunaan algoritma decision tree akan dilakukan percobaan pembagian rasio data yang berbeda - beda serta akan dibandingkan dengan algoritma klasifikasi lain. Hasilnya penerapan algoritma decision tree memiliki nilai akurasi terbaik sebesar 98.78% dengan rasio pembagian data 90:10 dan hasil model bisa diimplementasikan dalam bentuk aplikasi web sederhana. Penelitian ini diharapkan bisa dimanfaatakan dalam pemahaman, solusi dan alternatif yang lebih baik untuk pendeteksian web phishing.

Kata kunci: Deteksi Phishing, Decision Tree, API Virustotal, Data Mining.

ABSTRACT

In conventional phishing website detection, the most commonly used methods by antivirus vendors and researchers are blacklists, whitelists, and signatures, which still have many limitations, including their inability to detect new phishing websites or replica websites that have different content and other dangerous websites. Some researchers maximize their detection using data mining and machine learning algorithms and compete to achieve better accuracy. Therefore, this study proposes the use of machine learning and data mining with the decision tree algorithm for phishing website detection using data from the online VirusTotal API platform. The data obtained consists of URL status and the vendor that detected it, totaling 819 URLs. The decision tree algorithm will be experimented with different data ratio divisions and compared with other classification algorithms. The results showed that the implementation of the decision tree algorithm had the best accuracy value of 98.78% with a 90:10 data ratio division, and the model results can be implemented in the form of a simple web application. This research is expected to be utilized for better understanding, solutions, and alternatives for phishing website detection.

Keyword: Phishing Detection, Decision Tree, API Virustotal, Data Mining.