PERBANDINGAN LOGISTIC REGRESSION DAN RANDOM FOREST MENGGUNAKAN CORRELATION-BASED FEATURE SELECTION UNTUK DETEKSI WEBSITE PHISING

SKRIPSI

disusun oleh

Farida

17.11.1708

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2021

PERBANDINGAN LOGISTIC REGRESSION DAN RANDOM FOREST MENGGUNAKAN CORRELATION-BASED FEATURE SELECTION UNTUK DETEKSI WEBSITE PHISING

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika

disusun oleh

Farida

17.11.1708

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2021

PERSETUJUAN

SKRIPSI

PERBANDINGAN LOGISTIC REGRESSION DAN RANDOM FOREST MENGGUNAKAN CORRELATION-BASED FEATURE SELECTION UNTUK DETEKSI WEBSITE PHISING

yang dipersiapkan dan disusun oleh

Farida

17.11.1708

telah disetujuai oleh Dosen Pembimbing Skripsi

pada tanggal 29 Agustus 2021

Dosen Pembimbing,

Ali Mustopa, M.Kom.
NIK. 190302192

PENGESAHAN

SKRIPSI

PERBANDINGAN LOGISTIC REGRESSION DAN RANDOM FOREST MENGGUNAKAN CORRELATION-BASED FEATURE SELECTION UNTUK DETEKSI WEBSITE PHISING

yang dipersiapkan dan disusun oleh

Farida

17.11.1708

telah dipertahankan di depan Dewan Penguji
pada tanggal 18 November 2021

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Ali Mustopa, M.Kom **NIK. 190302192**	
Rizqi Sukma Kharisma, M.Kom **NIK. 190302215**	
Akhmad Dahlan, M.Kom **NIK. 190302174**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 08 Desember 2021

DEKAN FAKULTAS ILMU KOMPUTER

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

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Banjarnegara, 04 Desember 2021

Farida

NIM. 17.11.1708

MOTTO

"Waktu bagaikan pedang. Jika kamu tidak memanfaatkannya dengan baik, maka ia akan memanfaatkanmu."

(HR. Muslim)

PERSEMBAHAN

Alhamdulillaahi rabbil'aalamiin

Puji syukur kepada Allah SWT atas segala rahmat dan hidayah yang telah memberikan kelancaran, kemudahan, kesehatan, dan kesabaran dalam menyusun skripsi ini. Pada kesempatan ini penulis ingin mempersembahkan karya penelitian ini sekaligus mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua, Bapak Darno Nasrudin dan Ibu Parsiah, atas segala doa, dukungan dan kasih sayang yang tidak ada hentinya diberikan. Terima kasih atas segala pengorbanan, nasihat dan bantuan baik secara moril maupun materi hingga saat ini dan seterusnya.
2. Kedua adik saya, Rehana dan Dawam yang selalu menemani, memberikan doa dan semangat hingga penelitian skripsi ini dapat terselesaikan.
3. Untuk seluruh keluarga saya yang selalu memberikan doa dan dukungan.
4. Bapak Ali Mustopa, M.Kom selaku dosen pembimbing, terimakasih banyak atas bimbingannya sehingga skripsi ini dapat terselesaikan.
5. Untuk mas Rafi dan Anggun sebagai support system yang selalu ada dan membantu dalam proses skripsi ini sehingga mendapatkan hasil yang terbaik.

Seluruh pihak yang tidak dapat disebutkan satu per satu, terimakasih atas segala bantuannya dan do'anya sehingga terselesaikan skripsi ini.

KATA PENGANTAR

Assalamu'alaikum Warahmatullaahi Wabarakaatuh

Dengan rahmat Allah SWT Yang Maha Pengasih dan Penyayang, puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Perbandingan Logistic Regression Dan Random Forest Menggunakan Correlation-Based Feature Selection Untuk Deteksi Website Phising" dengan baik dan lancar.

Skripsi ini disusun sebagai salah satu syarat kelulusann program sarjana jurusan Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Penulis menyadari bahwa skripsi ini tidak mungkin terselesaikan tanpa aadanya dukungan, bantuan, bimbingan, dan nasihat dari bebagai pihak selama penyusunan skripsi ini. Pada kesempatan ini penulis menyampaikan terima kasih setulus-tulusnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Bapak Ali Mustopa, M.Kom selaku dosen pembimbing penulis yang selalu memberikan arahan dan bimbingan sehingga skripsi ini dapat diselesaikan dengan baik.
4. Bapak Rizqi Sukma Kharisma, M.Kom selaku dosen penguji yang telah memberikan kritik dan saran untuk perbaikan skripsi ini.
5. Bapak Akhmad Dahlan, M.Kom selaku dosen penguji yang telah memberikan masukan dalam skripsi ini.
6. Bapak Ibu dosen Universitas AMIKOM Yogyakarta yang telah memberikan bekal ilmu dan motivasi selama perkuliahan kepada penulis.
7. Para penulis sumber bacaan, jurnal, dan makalah yang penulis jadikan referensi dalam penulisan laporan skripsi ini.
8. Kedua orang tua, keluarga besar dan teman-teman yang senantiasa memberikan doa, dukungan dan semangat kepada penulis dalam menyelesaikan skripsi dan perkuliahan.

9. Sahabat-sahabatku ketika kuliah Anggun, Hafifah, Fay, Rizqi, Sulis, Tito, Rangga, Seyma, Fatimah dan semua anak IF12 2017 yang tidak bisa disebutkan semuanya.
10. Seluruh pihak yang telah membantu baik secara langsung maupun tidak langsung yang tidak bisa penulis sebutkan satu persatu.

Penulis mengucapkan terima kasih atas segala do'a dan dukungannya serta mohon maaf yang sebesar-besarnya. Semoga segala bantuan serta amal baik semua pihak diatas mendapat balasan yang setimpal dari Allah SWT. Penulis menyadari bahwa masih ada banyak kekurangan di dalam laporan ini. Namun penulis berharap laporan skripsi ini dapat memberikan manfaat kepada pembaca.

Akhir kata, penulis berharap laporan skripsi ini dapat berguna dan bermanfaat sebagai bahan kajian untuk mahasiswa yang akan melakukan penelitian selanjutnya.

Wassalaamu'alaikum warahmatullaahi wabarakaatuh

Banjarnegara, 04 Desember 2021

Penulis

Farida

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

INTISARI

Dunia pada saat ini tengah mengalami perkembangan pada bidang teknologi dan komunikasi secara massif, terlebih pada masa pandemi saat ini yang mengharuskan kita semua belajar bahkan bekerja secara daring. Hal ini yang memicu banyaknya kejahatan di dunia internet. Salah satunya yaitu mencuri data pengguna internet melalui sebuah website palsu yang dibangun seperti asli atau disebut juga website phising.

Pada penenitian ini, untuk mengatasi maraknya website phising di dunia maya, maka diperlukan suatu model klasifikasi untuk mendeteksi website yang terindikasi phising dengan menggunakan kinerja terbaik dari salah satu algoritma klasifikasi logistic regression dan random forest. Untuk meningkatkan kinerja algoritma klasifikasi dilakukan seleksi fitur menggunkan metode correlation-based feature selection (CFS) untuk menyeleksi atribut yang paling berpengaruh dalam mendeteksi web phising.

Berdasarkan hasil uji coba, penerapan algoritma klasifikasi logistic regression dan random forest dalam klasifikasi web phishing dihasilkan akurasi sebesar 93,035 % dan 96.834 %, setelah dilakukan seleksi fitur dengan CFS akurasi yang dihasilkan menjadi 92.718 % dan 97.015 %. Dari uji coba terjadi peningkatan akurasi pada random forest sebesar 0,181 % dan terjadi penurunan yang tidak signifikan pada logistic regression. Hasil uji coba membuktikan bahwa seleksi fitur dengan CFS dapat menghilangkan atribut redundan dan dihasilkan akurasi algoritma klasifikasi yang tidak jauh berbeda ketika atribut lengkap.

Kata Kunci: website phising, klasifikasi, *logistic regression*, *random forest*, *correlation-based feature selection* (CFS).

ABSTRACT

The world is currently experiencing mass developments of information technology, especially during the current pandemic which requires all of us to learn and even work online. This is what triggers a lot of crime in the internet world. One of them is stealing internet user data through a fake website that is built like the original or also called a phishing website.

In this research, in order to overcome the rise of phishing websites in cyberspace, a classification model is needed to detect phishing websites by using the best performance from one of the logistic regression and random forest classification algorithms. To improve classification performance, feature selection is carried out using the correlation-based feature selection (CFS) method to select the most influential attribute in detecting web phishing.

Based on the test results, the application of the logistic regression and random forest classification algorithm in the classification of web phishing resulted in an accuracy of 93.035% and 96.834%, after feature selection with CFS the resulting accuracy was 92.718% and 97.015%, respectively. From the trial, there was an increase in accuracy in random forest by 0.181% and an insignificant decrease in logistic regression. The test results prove that feature selection with CFS can eliminate redundant attributes and the resulting classification algorithm accuracy is not much different when the attributes are complete.

Keyword: *website phishing, classification, logistic regression, random forest, correlation-based feature selection (CFS).*