SELEKSI FITUR BERBASIS INFORMATION GAIN PADA KINERJA ALGORITMA KLASIFIKASI UNTUK DETEKSI WEB PHISHING

SKRIPSI

disusun oleh

Anggun Wahyu Andriyanto

17.11.1732

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

2021

SELEKSI FITUR BERBASIS INFORMATION GAIN PADA KINERJA ALGORITMA KLASIFIKASI UNTUK DETEKSI WEB PHISHING

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika

disusun oleh

Anggun Wahyu Andriyanto

17.11.1732

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021

PERSETUJUAN

SKRIPSI

SELEKSI FITUR BERBASIS INFORMATION GAIN PADA KINERJA ALGORITMA KLASIFIKASI UNTUK DETEKSI WEB PHISHING

yang dipersiapkan dan disusun oleh

Anggun Wahyu Andriyanto
17.11.1732

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 04 Januari 2021

Dosen Pembimbing,

Erni Seniwati, S.Kom., M.Cs.
NIK. 190302231

PENGESAHAN

SKRIPSI

SELEKSI FITUR BERBASIS INFORMATION GAIN PADA KINERJA ALGORITMA KLASIFIKASI UNTUK DETEKSI WEB PHISHING

yang dipersiapkan dan disusun oleh

Anggun Wahyu Andriyanto

17.11.1732

telah dipertahankan di depan Dewan Penguji
pada tanggal 15 Januari 2021

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Erni Seniwati, S.Kom, M.Cs **NIK. 190302231**	
Lilis Dwi Farida, S.Kom, M.Eng **NIK. 190302288**	
Irma Rofni Wulandari, S.Pd, M.Eng **NIK. 190302329**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Januari 2021

DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si, M.T
NIK. 190302038

v

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.

Boyolali, 2 Februari 2021

Anugun Wahyu Andriyanto

NIM. 17.11.1732

MOTTO

"Apa adanya dan adanya apa"

(Tara Arts)

PERSEMBAHAN

Puji syukur penulis panjatkan kepada Allah SWT, yang telah memberikan kesehatan, rahmat dan hidayah sehingga penulis diberikan kesempatan untuk menyelesaikan skripsi ini, sebagai salah satu syarat untuk mendapatkan gelar kesarjanaan. Walaupun jauh dari kata sempurna, namun penulis bangga telah mencapai pada titik ini, yang akhirnya skripsi ini bisa selesai diwaktu yang tepat. Skripsi ini saya persembahkan untuk :

1. Bapak dan Ibu, terima kasih atas doa, semangat, pengorbanan, serta kasih saying yang tidak pernah henti sampai saat ini.
2. Adikku Putri, terima kasih telah menjadi penyemangat dan tempat keluh kesah dalam proses pengerjaan skripsi.
3. Bapak Yoga Pristyanto, terima kasih telah memberikan ide dan gambaran dalam menentukan topik skripsi.
4. Ibu Erni Seniwati yang telah sabar membimbing saya karena sering menghilang tanpa kabar.
5. Warung Barokah atau yang sering disebut Mbok darmi, terima kasih karena dari PPM sampai mengambil skripsi selalu menjadi penyelamat perut dikala kelaparan dengan harga yang membumi.
6. Teman seperjuangan Rizqi dan Sulis dalam kelompok pendeteksi web phishing.
7. Buaji, Gita Mama, Nanda Okto, Mas Dayat, Yoga, Sella, Fery, Novita, Bahrul yang sudah membantu, memberikan masukan dan semangat dalam mengerjakan skripsi ini.
8. Semua teman-teman Informatika 12.
9. Kepada teman-teman, saudara dan dosen yang tidak bisa saya sebutkan satu persatu.

KATA PENGANTAR

Assalamualaikum Wr. Wb.

Dengan rahmat Allah SWT Yang Maha Pengasih dan Penyayang, puji syukur penulis panjatkan kehadirat Allah SWT yang telah memberikan dan menganugrahkan kasih sayang, rezeki, dan kesehatan serta atas berkah, ridho dan hidayah-Nya, sehingga saya sebagai penulis dapat menyelesaikan skripsi dengan judul "Seleksi Fitur Berbasis Information Gain Pada Kinerja Algoritma Klasifikasi Untuk Deteksi Web Phishing". Penyusunan skripsi ini dibuat sebagai salah satu syarat kelulusan program sarjana jurusan Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta. Penulis menyadari selama menempuh pendidikan dan proses penyelesaian skripsi ini memperoleh bantuan dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan kali ini penulis ingin mengucapkan rasa terima kasih yang mendalam dan tak terkira kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku rektor Universitas AMIKOM Yogyakarta.
2. Ibu Krisnawati, S.Si, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Bapak Sudarmawan, M.T. selaku Ketua Program Studi Informatika Universitas AMIKOM Yogyakarta.
4. Ibu Windha Mega Pradnya D, M. Kom. selaku Sekretaris Program Srudi Informatika Universitas AMIKOM Yogyakarta.
5. Ibu Erni Seniwati, S.Kom, M.Cs. selaku dosen pembimbing, terima kasih banyak atas bimbingan yang telah diberikan dalam membimbing penulis sehingga skripsi ini dapat diselesaikan dengan baik.
6. Ibu Lilis Dwi Farida, S.Kom, M.Eng. selaku dosen penguji yang telah banyak memberikan masukan dan bimbingan dalam skripsi ini.
7. Ibu Irma Rofni Wulandari, S.Pd, M.Eng. selaku dosen penguji, terima kasih banyak atas bimbingan, kritik dan sarannya untuk perbaikan skripsi ini.
8. Seluruh Bapak dan Ibu dosen Fakultas Ilmu Komputer Jurusan Informatika yang telah memberikan bekal ilmu kepada penulis.

9. Bapak Yoga Pristyanto, S.Kom, M.Eng. selaku dosen mata kuliah Kecerdasan Buatan yang telah memberikan ide sebagai topik pada skripsi ini.
10. Seluruh staff dan karyawan Universitas AMIKOM Yogyakarta.

Bagi seluruh pihak yang tidak bisa penulis sebutkan namanya satu persatu, penulis mengucapkan rasa terima kasih atas segala doa dan dukungannya serta mohon maaf yang sebesar-bersarnya. Semoga segala kebaikan, bantuan dana amal baik dari berbagai pihak tersebut diatas mendapat balasan yang setimpal dari Allah SWT dan penulis senantiasa berharap semoga skripsi yang dibuat in dapat bermanfaat untuk berbagai pihak.

Wassalamualaikum Wr. Wb.

Boyolali, 25 Februari 2021

Penulis,

Anggun Wahyu Andriyanto

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

INTISARI

Pencurian data pengguna dari internet melalui web palsu yang dirancang menyerupai asli disebut web phishing. Web phishing juga terjadi pada media sosial dan internet *banking*. Web phishing tersebut akan mengelabuhi pengguna internet untuk memasukkan informasi penting seperti *password* ataupun data bank, sehingga pelaku phishing bisa mendapatkan data tersebut.

Pada penelitian ini, untuk mencegah kriminalitas phishing sehingga mengurangi dan bisa menghindari kerugian situs phishing terhadap pengguna internet maka dibutuhkan suatu model klasifikasi untuk memprediksi web yang terindikasi phishing, menggunakan kinerja terbaik dari salah satu algoritma klasifikasi *decision tree* (CART), k-*nearest neighbor*, dan *naïve bayes*. Untuk meningkatkan kinerja dari algoritma klasifikasi maka dilakukan seleksi fitur menggunakan metode *information gain* untuk menyeleksi atribut yang paling berpengaruh dalam mendeteksi web phishing.

Berdasarkan hasil uji coba, penerapan algoritma klasifikasi *decision tree* (CART), k-*nearest neighbor*, dan *naïve bayes* dalam klasifikasi web phishing dihasilkan akurasi sebesar 96,56%, 95,07%, dan 90,14%, setelah dilakukan seleksi fitur dengan metode *information gain* akurasi yang dihasilkan 97,01%, 94,84%, dan 90,09%. Dari uji coba terjadi peningkatan akurasi pada *decision tree* (CART), penurunan yang tidak signifikan bahkan cenderung stabil pada k-*nearest neighbor* dan *naïve bayes* setelah dihilangkannya beberapa fitur. Dari hasil uji coba terbukti bahwa seleksi fitur dengan metode *information gain* mampu menghilangkan atribut redundan, lalu dihasilkan akurasi algoritma klasifikasi yang tidak jauh berbeda dengan data ketika atributnya masih lengkap dan bias diterapkan untuk mendeteksi web phishing.

Kata Kunci: web phishing, klasifikasi, deteksi phishing, *decision tree* (CART), k-*nearest neighbor*, *naïve bayes*, *information gain*.

ABSTRACT

The theft of user data from the internet via a fake web designed to resemble the real thing or called web phishing. Web phishing also occurs in social media and internet banking. The web phishing will trick internet users into entering important information such as passwords or bank data, so that phisers can get the data.

In this study, to prevent phishing crime so as to reduce and avoid loss of phishing sites to internet users, a classification model is needed to predict phishing-indicated webs, using the best performance from one of the decision tree (CART) classification algorithms, k-nearest neighbor, and naïve bayes. To improve the performance of the classification algorithm, feature selection is carried out using the information gain method to select the most influential attributes in detecting web phishing.

Based on the test results, the application of the decision tree (CART), k-nearest neighbor, and naïve bayes classification algorithm in the web phishing classification resulted in an accuracy of 96.56%, 95.07%, and 90.14%, after selecting features with the information gain method resulted in the accuracy of 97.01%, 94.84%, and 90.09%. From the testing, there was an increase in the accuracy of the decision tree (CART), the decrease was not significant and even tended to be stable in the k-nearest neighbor and naïve bayes after removing some features. From the test results, it is proven that feature selection using the information gain method is able to eliminate redundant attributes and the resulting classification algorithm accuracy is not much different when the attributes are complete.

Keyword: *website phishing, classification, phishing detection, decision tree (CART), k-nearest neighbor, naïve bayes, information gain.*