

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini :

Nama : Aswan Supriyadi Sunge, S.E, M.Kom
Tempat/Tgl Lahir : Jakarta, 26 Januari 1980
Alamat : Perum Mutiara Bekasi Jaya Blok E8. 10 Cibarusah
Sindang Mulya Kabupaten Bekasi
E-Mail : aswan.sunge@pelitabangsa.ac.id
Telp/Handphone : 082299218720

Dengan ini menyatakan bahwa tulisan yang telah saya buat dengan judul:

"Optimasi Algoritma C4.5 Dalam Prediksi Web Phishing Menggunakan Seleksi Fitur Genetic Algoritma"

adalah hasil tulisan sendiri, dan semua sumber baik yang kutip maupun yang dirujuk telah saya nyatakan dengan benar dan tulisan belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tulisan yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan tulisan dianggap batal.

Jakarta, 15 Agustus 2018
Yang menyatakan,



(Aswan Supriyadi Sunge)



Plagiarism Checker X Originality Report

Similarity Found: 16%

Date: Thursday, August 16, 2018

Statistics: 322 words Plagiarized / 2013 Total words

Remarks: Low Plagiarism Detected - Your Document needs Optional Improvement.

Optimasi Algoritma C4.5 Dalam Prediksi Web Phishing Menggunakan Seleksi Fitur Genetic Algoritma Aswan Supriyadi Sunge Sekolah Tinggi Teknologi Pelita Bangsa e-mail: aswan.sunge@pelitabangsa.ac.id Cara Sitasi: diisi oleh editor Abstract - Salah satu isu terpenting saat ini dalam dunia online yaitu keamanan.

Masalah keamanan terbesar salah satunya adalah Phishing yang melibatkan duplikat situs yang sah atau asli untuk menipu dengan mencuri informasi pengguna online. Memang diakui sangat sukar untuk membedakan situs asli dengan palsu. Oleh sebab itu dibutuhkan klasifikasi dalam memprediksi website yang terindikasi Phishing. Dengan klasifikasi dalam Algoritma C4.5, permasalahan tersebut dapat diselesaikan dengan menghasilkan rule dari pohon keputusan.

Untuk dapat meningkatkan akurasi dari prediksi algoritma C4.5 dapat digunakan fitur seleksi dengan menggunakan algoritma genetika. Berdasarkan penerapan algoritma C4.5 dihasilkan akurasi sebesar 83,81% untuk memprediksi website Phishing dan dengan seleksi fitur menggunakan algoritma genetika meningkatkan akurasi sebesar 3,22% menjadi 86,47. Dari penelitian ini algoritma genetika terbukti dapat meningkatkan akurasi untuk prediksi website phishing.. Keywords: phishing, prediksi, algoritma C4.5, algoritma genetika PENDAHULUAN Perkembangan internet begitu signifikan, jika dilihat pertumbuhan internet di dunia sudah lebih dari 4 milyar dan di Indonesia lebih dari 143 juta pengguna (<https://internetworldstats.com/stats>, 2018) dari total populasi lebih dari 266 juta penduduk.

Hal ini disebabkan berkembang teknologi informasi yang sangat cepat dan berbagai macam media dan fungsi yang salah satunya dalam hal transaksi keuangan maupun e-commerce. Hal tersebut memudahkan pelanggan tanpa harus bersusah payah dan

tanpa perlu keluar rumah. Tetapi di dalam kemudahan bertransaksi muncul salah satu masalah terbesar yaitu keamanan bertransaksi.

Ini menjadi momok menakutkan bagi pengguna online, apalagi sudah merambah dalam pengguna social media (Wibowo, Mia & Fatiman, 2017). Satu hal dari keamanan dari ketidaktahuan dari segi pengguna yang akibatnya terjerumus ke dunia Cybercrime. Juga banyak pengguna online tidak bisa membedakan antara situs asli maupun situs palsu atau Phishing, maka dari itu penelitian ini bertujuan untuk bisa memprediksi situs yang terindikasi Phishing..

TINJAUAN PUSTAKA Phishing merupakan metode ataupun cara dalam mengelabui pengguna online dan paling umum dalam serangannya dengan memberikan link atau pesan email ke situs yang tampaknya asli (Junaind, Shafique, Robert, 2016). Teknik pun semakin beragam bukan hanya membuat situs asli atau memberikan link tapi menggunakan mobile (Belal, Amro, 2018), inilah sebagai celah dalam ketidak tahun pengguna online. Memang diakui metode maupun cara sulit dalam mendeteksi apalagi seorang pengguna yang tidak tahu akan keamanan.

Maka dari itu dibutuhkan prediksi dalam mendeteksi terindikasi Phishing, untuk itu dibutuhkan klasifikasi dalam data mining (Mofleh, Al-diabat, 2016) dalam melihat data maupun parameter yang ada yang dijadikan patokan dalam pendekteksian Phishing.. Data mining merupakan asal kata dari mining yang berarti tambang, dikembangkan menjadi konsep dalam melihat informasi maupun pengetahuan, dari data lampau maupun masa lalu yang tersimpan dalam database (Larose, 2005) dan penggunaan data mining digunakan untuk menganalisis suatu perilaku maupun prediksi, juga bukan hanya digunakan dalam ilmu computer saja tetapi bidang lain seperti bisnis maupun industri (Giudici & Figini. 2009).

Istilah data mining maupun Knowledge Discovery in Databases (KDD) tidak lepas dari keduanya dikarenakan menggali data yang tersimpan dalam data yang sangat besar (Fayyad, U.; Piatetsky-Shapiro, G; Smyth, 1996). Skema tersebut digambarkan dibawah ini, Sumber (Fayyad, U.; Piatetsky-Shapiro, G; Smyth, 1996) Gambar 1. Proses Skema KDD Tahapan proses KDD dalam data mining sebagai berikut : 1. Data Selection, data akan diseleksi berdasarkan kecocokan data yang akan diambil keputusan.

2. Data Preprocessing/Cleaning,, pada tahap ini dilakukan pembersihan data yang kosong, duplicate atau tidak sesuai dari yang ingin diputuskan 3. Transformation, melihat data yang sudah dipilih dan dipresentasikan dari hasil yang diinginkan. 4. Data Mining, melihat pola yang ingin ditampilkan dari metode maupun Teknik yang dipilih sebelumnya misalnya klasifikasi, clustering, regresi, CART dan lain sebagainya.

5. Interpretation/Evaluation, pada tahap ini sebagai penerjemah dari data yang telah ditampilkan dan melihat hasil dari teknik atau metode yang digunakan. Algoritma **Algoritma dalam klasifikasi yang** banyak digunakan ialah Decision Tree. Dikarenakan sangat mudah dimengerti dan dijabarkan oleh banyak pengguna juga mudah dipahami dimana cabang pohon disimpulkan dalam bentuk klasifikasi (Gorunescu.2011).

Pohon **keputusan mempunyai tiga pendekatan klasik** ; 1. Classification, melihat hasil prediksi berdasarkan kelas atau label(misalnya: Ya atau Tidak, Mati atau Tidak Mati). 2. Regression, melihat hasil prediksi belum tentu akan hasilnya(misalnya : Pemberian Kredit, Pencapaian Target Produksi) 3.

CART (Classification & Regression Tree), yaitu berdasarkan susunan pertanyaan yang saling berkaitan dan berurutan dan hasil jawaban tersebut menjabarkan pertanyaan berikutnya. Namun jika pertanyaan tidak sesuai maka akan berhenti dan tidak melanjutkan pertanyaan. Dari setiap pohon keputusan menghasilkan simpul yang merupakan hasil prediksi atau solusi untuk menghasilkan solusi dari pertanyaan yang saling berkaitan (Seemam Rathi, Mamta, 2012). Sumber (Seemam Rathi, Mamta, 2012) Gambar 2.

Ilustrasi Decision Tree Algoritma dalam Decision Tree banyak sekali (Wu, Xindong, 2007) namun yang banyak digunakan yaitu ID3 dan Algoritma C4.5. Kedua mempunyai kesamaan **dikarenakan Algoritma C4.5 merupakan pengembangan dari ID3** namun ada perbedaan yang utama yaitu : a. Ketika data (atribut) yang berkelanjutan atau putus-putus terutama berhubungan data traninig maka Algoritma C4.5 dapat memperbaikinya. b.

Hasil yang didapat dari Algoritma C4.5 dapat dipangkas ketika terbentuk c. Penyeleksian variabel dilakukan dengan Gain Ratio. Perubahan dari ID3 ke C4.5 dalam **Gain Ratio** untuk diperbaharui information gain maka dengan rumus :(1) Keterangan: $S =$ Ruang/Data sample yang dipergunakan untuk data training $A =$ Atribut $Gain(S,A) =$ Information gain pada atribut A $SplitInfo(S,A) =$ split information pada atribut A Pemilihan atribut dari Gain Ratio yang tertinggi dijadikan **sebagai atribut test untuk simpul.**

Pendekatan ini menerapkan konsep **normalisasi pada information gain** yang disebut dengan **split information** dengan rumus dibawah ini :(2) Keterangan : $S =$ Ruang (data) sample yang digunakan untuk training. $A =$ Atribut. $S_i =$ Jumlah sample untuk atribut i Pada tahun 1970 Algoritma Genetika (GA) diperkenalkan oleh John Holland di Universitas Michigan (J.H).

Holland, 1975), bahwa dari bagian masalah merupakan bentuk dari adaptasi dari alam maupun buatan yang dapat diformulasikan menjadi bagian genetika (Suryanto, 2007). GA merupakan bagian optimasi dan **pencarian yang didasarkan pada seleksi alam** dan seleksi makhluk hidup secara apa adanya. Pada akhirnya, mengembalikan satu bagian yang terbaik yang dijadikan **solusi dari masalah yang akan** dipecahkan sebagai kromosom (Desiani, A.,

& Muhammad, A, 2006) Ada tiga aspek dalam dalam menggunakan GA : 1. Definisi fungsi objektif/definisi 2. Definisi dan implementasi representasi genetika. 3. Definisi dan implementasi dari operator genetika. METODOLOGI PENELITIAN Sample dalam penelitian ini merupakan ini merupakan data sekunder yang didapat dari hasil komputasi digital pada UCI Neda Abdelhamid Auckland Institute of Studies. Data yang didapat terdiri dari Variabel Rendah (0), Sedang (-1) dan Tinggi (1).

Untuk paramaternya terdapat 9 yaitu SFH, popUpWindow ,SSLfinal_State, Request_URL, URL_of_Anchor, web_traffic, URL_Length, age_of_domain, having_IP_Address. Keseluruhan data berjumlah 1353 data kemudian dibagi 2 bagian yang dijadikan data training maka akan diperoleh decision tree untuk hasil klasifikasi berjumlah 1081 data dan data testing untuk melihat akurasi dari klasifikasi tersebut berjumlah 272 data. **Untuk mengukur tingkat akurasi** dari prediksi menggunakan Rapid Miner.

Tahap ini dalam penelitian sebagai berikut : 1. Pengumpulan (pengambilan) data Pada tahap ini mencari data yang tersedia, mengintegrasikan semua data kedalam data set, termasuk variabel yang diperlukan dalam proses 2. Pengolahan data awal Ditahap ini dilakukan penyeleksian, pembersihan termasuk melihat data yang kosong kemudian merubah data yang diinginkan. 3.

Metode **yang diusulkan Pada tahap ini** penganalisisan data kemudian pengelompokan variabel yang saling berhubungan dengan yang lain, kemudian penerapan model yang sesuai data yang telah dibentuk. 4. **Eksperimen dan pengujian metode** Pada tahap ini penentuan model yang diusulkan ketika akan diuji dan melihat hasil rules yang dijadikan pengambilan keputusan. 5.

Evaluasi dan validasi Pada tahap ini melakukan hasil evaluasi yang didapat dari model yang ditetapkan sebelum dan melihat hasil akurasi dengan pengujian aplikasi terhadap metode yang digunakan. Dibawah ini gambar skema dalam tahapan penelitian yang dilakukan : Gambar 3. Kerangka Pemikiran HASIL DAN PEMBAHASAN Dari hasil data yang diperoleh kemudian dikategorikan dengan variabel yang ada dengan nilai rendah, sedang dan tinggi dan pengujian berdasarkan **data training dan data** testing. Table 1.

Data Training Terbentuk rule yang diperoleh dengan pengujian dengan Rapidminer didapat decision tree dalam prediksi website Phishing. Sumber : Rapidminer Gambar 4. Rule Decision Tree Setelah didapat rule decision tree kemudian melihat confusion matrix. Akurasi dari data training dari 1082 data dihasilkan accuracy sebesar 83.81% Table 2.

Akurasi Data Training Akurasi dari data testing yang berjumlah 272 data dihasilkan accuracy sebesar 81.94%. Table 3. Akurasi Data Testing Setelah melihat akurasi dari data training dan testing pada Algoritma C4.5 kemudian di optimasi hasilnya dengan Algoritma Genetika. Dari akurasi yang didapat dengan fitur seleksi algoritma genetika sebesar 86.40%. Table 4. Akurasi Data Training dengan GA Table 5.

Akurasi Data Testing dengan GA sebesar 88.58% Jika dilihat perbandingan antara Algoritma C4.5 dengan Algoritma Genetika. Data Metode Algoritma C4.5 Algoritma Genetika Training 83.81% 86.40% Testing 81.94% 88.58% KESIMPULAN Dari hasil pembahasan diambil kesimpulan sebagai berikut : 1. Algoritma C4.5 dengan Fitur Seleksi Algoritma Genetika maka indikasi website Phishing dapat diprediksi dan dapat dijadikan kontribusi terhadap proses pengambilan keputusan ke pengguna online. 2. Evaluasi dalam pengujian hasil prediksi dari Decision Tree algoritma C4.5

dengan seleksi fitur algoritma genetika, dan hasil prediksi yang didapatkan dalam pengujian ini adalah 86,40% hasil ini meningkat dari penelitian yang sebelumnya menggunakan data yang sama dan algoritma yang sama yaitu algoritma decision tree hasil prediksinya adalah 83,81%, sehingga dapat disimpulkan bahwa tingkat dengan penggunaan seleksi fitur algoritma genetika mendapatkan hasil yang lebih baik dengan tingkat akurasi yang meningkat.

Berdasarkan hasil penelitian yang didapat memberikan saran sebagai berikut : 1. Perlu penelitian lebih lanjut dengan melakukan pengujian dengan metode lain maupun dibandingkan seperti SVM, k-NN, Neural Network, Naïve Bayes dan lain-lain agar melihat hasil perbandingan dengan akurasi yang tertinggi dalam prediksi yang terindikasi website Phishing. 2.

Perlu diterapkan lebih lanjut optimasi menggunakan metode lain seperti Adaboost, atau PSO untuk mengetahui peningkatan akurasi dengan seleksi fitur. REFERENSI Amro, Belal, (2018). Phishing Techniques in Mobile Devices, Journal of Computer and Communications, , 6, 27-35 Al-Diabat, M. (2016). Detection and Prediction of Phishing Websites using Classification Mining Techniques. International Journal of Computer Applications, 147(5), 975-8887.

Retrieved from

<https://pdfs.semanticscholar.org/2c0b/81243b5011ed040ce31a4c2c48d2ba181ce2.pdf>

Chaudhry Junaid, Chaudhry Shafique, Rittenhouse Robert, (2016). Phishing Attacks and Defenses, *Internasional Journal of Security and its Applications*, V ol. 10, No. 1 (2016), pp.247-256
Desiani, A., & Muhammad, A., (2006). *Konsep Kecerdasan Buatan*. Yogyakarta: Cv. Andi Offset.
Fayyad, U.; Piatetsky-Shapiro, G; Smyth, P, (1996).

From *Data Mining to Knowledge Discovery: An overview in Advances in Knowledge discovery and Data Mining*; Uthurusamy, R. MIT Press. Cambridge, Mass.. pp. 1-36
Giudici & Figini. (2009). *Applied Data Mining for Business and Industry*, 2nd Edition.
Gorunescu.(2011). *Data Mining Concepts, Models and Techniques*. Romania.
Springer-Verlag Berlin Heidelberg.
J.H. Holland, (1975). *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, Ann Arbor, MI.

Larose, (2005), "Discovering Knowledge in Data: An Introduction to Data Mining", John Willey & Sons, Inc.
Mia Haryati Wibowo dan Nur Fatimah, (2007). *Ancaman Phishing Terhadap Pengguna Sosial Media Dalam Dunai Cyber Crime* Volume 1 Nomor 1 : 1 - 5.
Seema, Rathi Monika, Mamta, (2013). *Decision Tree: Data Mining Techniques*, *International Journal of Latest Trends in Engineering and Technology* (IJLTET). Suryanto. (2007).

Artificial Intelligent, Searching, Reasoning Planning dan Learning. Bandung: Informatika Bandung.
Wu, Xindong, (2007) "Top 10 Algorithms in Data Mining", Received: 9 July 2007 / Revised: 28 September 2007 / Accepted: 8 October 2007.
Vercellis, Carlo. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. United Kingdom: John Willey & Son <http://www.internetworldstats.com/stats> (2018)
Aswan Supriyadi Sunge, S.E, M.Kom, lahir di Jakarta, 26 Januari 1980. Sebagai penulis beberapa buku komputer dan praktisi VB.Net juga sebagai staff dan pengajar di Sekolah Tinggi Teknologi Pelita Bangsa sejak tahun 2014.

Menyelesaikan Studi S2 di Pascasarjana STMIK Nusa Mandiri Jakarta program studi Ilmu Komputer. Penelitian yang pernah dilakukan seperti : (1) *Komparasi Menggunakan Algoritma C4.5, Neural Network dan Naïve Bayes* Dalam Prediksi Ujian Kompetensi SMK Mahadhika 4 Jakarta, Terbit di Seminar Nasional Ilmu Pengetahuan dan Teknologi Komputer 2 (1), 391-397 Vol. 2014. (2) *Prediksi Ujian Kompetensi Dengan Menggunakan Klasifikasi Algoritma C4*.

5 Di SMK Mahadhika 4 Jakarta, terbit di *Bina Insani ICT Journal* 1 (2), 136-150 Vol. , 2014.
(3) *Prediksi Kompetensi Karyawan Menggunakan Algoritma C4.5 (Studi Kasus : PT Hankook Tire Indonesia)* terbit di *Seminar Nasional Teknologi Informasi dan Komunikasi*

Universitas Atmajaya Yogyakarta tanggal 23 -24 Maret 2018. Nomor ISSN Publikasi Online Sentika : 2337-3377. (4) Optimasi Algoritma C4.5

Menggunakan Genetic Algoritma Dalam Memprediksi Website Phishing terbit di Seminar Nasional Inovasi dan Tren (SNIT) 2018 di Bina Sarana Informatika Jakarta tanggal 25 Juli 2018 Prosiding SNIT 2018 Vol 1, No.1 ISBN:978-602-61268-5-6 n.

INTERNET SOURCES:

1% - <http://elibrary.nusamandiri.ac.id/listbuku/tesis>
<1% - http://eprints.dinus.ac.id/15306/1/jurnal_15343.pdf
1% -
<https://www.scribd.com/document/359555843/KOMPARASI-PEMODELAN-DATA-C4-5-PSO>
<1% - <https://id.m.wikipedia.org/wiki/UK>
<1% -
<https://www.scribd.com/doc/51892413/DAMPAK-TEKNOLOGI-INFORMASI-DALAM-BISNIS>
<1% - <https://www.hipwee.com/list/cara-memulai-bisnis-online-dari-rumah/>
<1% - <http://bbsnews.co.id/jaman-dianggap-tak-mampu-atasi-kemacetan/>
<1% - <http://niketutbudiartini.blogspot.com/2011/>
<1% - <https://nonosun.wordpress.com/2011/05/12/sia-2-10-database/>
<1% -
<https://www.scribd.com/document/235948168/Marselina-Prediksi-Kelulusan-Dengan-Naive-Bayes-Dan-C45-2010>
<1% - <https://www.scribd.com/document/182238253/TESIS>
<1% -
<https://docplayer.info/31321236-Prosiding-seminar-nasional-vokasi-dan-teknologi-ke-1-semnasvoktek.html>
3% - <https://fti.uajy.ac.id/sentika/publikasi/makalah/2018/3.pdf>
<1% -
<http://repository.ipb.ac.id/bitstream/handle/123456789/46235/g06syp.pdf?sequence=1>
1% -
<https://pemrogramanmatlab.com/data-mining-menggunakan-matlab/pohon-keputusan-menggunakan-matlab/>
<1% - <https://es.scribd.com/doc/39316488/Prodi-Teknik-Informatika>
1% - <https://informatikalogi.com/algoritma-c4-5/>
<1% -
<http://docplayer.info/29599028-Bab-ii-landasan-teori-digunakan-sebagai-pedoman-perawatan-adalah-sebuah-panduan-sebagaimana.html>

<1% - <https://dekasasmita.wordpress.com/2014/07/05/genetic-algorithm/>

<1% - <http://docplayer.info/45727151-Vol-iii-no-1-februari-2017-issn.html>

<1% - <http://docplayer.info/47728698-Penentuan-penilaian-kredit-menggunakan-metode-naive-bayes-berbasis-particle-swarm-optimization.html>

<1% - <https://docplayer.info/68167376-Skripsi-prediksi-hasil-pemilihan-umum-legislatif-dprd-provinsi-jawa-tengah-menggunakan-metode-decision-tree-dan-algoritma-c4-5.html>

<1% - https://www.researchgate.net/publication/320403762_Pemodelan_dan_Prototipe_Sistem_Informasi_untuk_Prediksi_Pembaharuan_Polis_Asuransi_Mobil_Menggunakan_Algoritma_C45

<1% - <https://www.scribd.com/doc/304852922/Jurnal-Data-Mining>

<1% - <https://www.scribd.com/document/339428120/Teknologi-Data-Mining-untuk-Prediksi-Laj-doc>

<1% - http://eprints.dinus.ac.id/16653/1/jurnal_15660.pdf

<1% - <https://www.scribd.com/document/334672200/IMPLEMENTASI-ALGORITMA-GENETIKA-UNTUK-MINIMASI-GALAT-PADA-METODE-PERAMALAN-ARIMA>

1% - <http://blogkumani.blogspot.com/2013/11/makalah.html>

1% - <https://pdfs.semanticscholar.org/2c0b/81243b5011ed040ce31a4c2c48d2ba181ce2.pdf>

1% - <https://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/KDD3.htm>

1% - https://link.springer.com/chapter/10.1007%2F978-3-662-04448-3_10

<1% - <https://repository.widyatama.ac.id/xmlui/bitstream/handle/123456789/2362/daftar%20pustaka.pdf?sequence=13>

1% - <http://repository.usu.ac.id/bitstream/handle/123456789/42373/Reference.pdf;sequence=2>

1% - <http://egan-jy.blogspot.com/2016/06/pengertian-jst-jaringan-syaraf-tiruan.html>

<1% - <https://dl.acm.org/citation.cfm?id=2953334>

<1% - <http://digilib.esaunggul.ac.id/public/UEU-Undergraduate-1119-Daftar%20Pustaka.pdf?ocezwjclgdgtdqesi>

1% - <http://ejournal.nusamandiri.ac.id/ejurnal/index.php/pilar/article/viewFile/343/275>

<1% - <http://docplayer.info/34641481-Kajian-penerapan-algoritmac4-5-neural-network-dan-naive-bayes-untuk-klasifikasi-mahasiswa-yang-bermasalah-dalam-registrasi.html>

<1% - <http://seminar.bsi.ac.id/snit/index.php/snit-2018>