

**OPTIMASI ALGORITMA KLASIFIKASI *DECISION TREE* (CART)  
DENGAN METODE *BAGGING* UNTUK  
DETEKSI *WEBSITE PHISHING***

**SKRIPSI**



disusun oleh

**Rizqi Krisandika**

**17.11.1724**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2020**

**OPTIMASI ALGORITMA KLASIFIKASI *DECISION TREE* (CART)  
DENGAN METODE *BAGGING* UNTUK  
DETEKSI *WEBSITE PHISHING***

**SKRIPSI**

untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh

**Rizqi Krisandika**

**17.11.1724**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2020**

# PERSETUJUAN

## SKRIPSI

### OPTIMASI ALGORITMA KLASIFIKASI *DECISION TREE* (CART) DENGAN METODE *BAGGING* UNTUK DETEKSI *WEBSITE PHISHING*

yang dipersiapkan dan disusun oleh

**Rizqi Krisandika**

17.11.1724

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 29 September 2020

**Dosen Pembimbing,**

**Arif Akbarul Huda, S.Si., M.Eng.**

**NIK. 190302287**

**PENGESAHAN**

**SKRIPSI**

**OPTIMASI ALGORITMA KLASIFIKASI *DECISION TREE* (CART)  
DENGAN METODE *BAGGING* UNTUK  
DETEKSI *WEBSITE PHISHING***

yang dipersiapkan dan disusun oleh

**Rizqi Krisandika**

**17.11.1724**

telah dipertahankan di depan Dewan Penguji  
pada tanggal **21 September 2020**

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Dony Ariyus, M.Kom**  
**NIK. 190302128**

**Theopilus Bayu Sasongko, S.Kom, M.Eng**  
**NIK. 190302375**

**Arif Akbarul Huda, S.Si, M.Eng**  
**NIK. 190302287**

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 26 September 2020

**DEKAN FAKULTAS ILMU KOMPUTER**

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

## 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.

Yogyakarta, 1 Oktober 2020



Rizqi Krisandika

NIM. 17.11.1724

## **MOTTO**

“Karena sesungguhnya sesudah kesulitan itu ada kemudahan. Sesungguhnya sesudah kesulitan itu ada kemudahan”

**(Q.S. Al-Insyirah 5-6)**

”Usaha keras itu tak akan mengkhianati”

**(Shonichi - JKT48)**



## PERSEMBAHAN

Puji syukur kepada Allah SWT atas segala rahmat dan hidayah yang telah memberikan kelancaran, kemudahan, kesehatan, dan kesabaran dalam menyusun skripsi ini. Skripsi ini saya persembahkan kepada orang-orang yang saya sayangi dan cintai :

Kedua orang tua dan adik saya, Bapak Fajar Krisdianto, Ibu Neneng Tati Dewi, dan Rizqi Alnova Defario yang tidak pernah lelah memberikan doa, dukungan, motivasi, dan finansial serta telah mengajarkan untuk selalu berusaha dan sabar dalam menjalani kerasnya hidup. Terima kasih telah memberikan yang terbaik untuk saya selama ini atas jerih payah yang dilakukan agar dapat melanjutkan kuliah hingga lulus sebagai sarjana. Semoga segala sesuatu yang saya lakukan dapat membahagiakan keluarga terutama kedua orang tua. Semoga ayah dan mama selalu diberikan kesehatan dan kebahagiaan hingga saya sukses nantinya.

## KATA PENGANTAR

Assalamu'alaikum Wr. Wb

Puji syukur penulis panjatkan kepada Allah SWT yang telah memberikan rahmat, karunia, dan hidayahnya dalam penulisan skripsi ini sehingga penulis dapat menyelesaikannya dengan baik dan lancar. Sholawat serta salam senantiasa dicurahkan kepada junjungan Nabi Muhammad SAW sebagai suri tauladan bagi seluruh umat manusia dan pengikutnya hingga akhir zaman, Aamiin.

Laporan skripsi ini disusun sebagai salah satu syarat kelulusann program S1-Informatika Universitas Amikom Yogyakarta. Dalam penyusunan laporan skripsi ini penulis mendapatkan bantuan dari berbagai pihak yang telah membantu dalam penulisan laporan skripsi ini. Maka dari itu penulis mengucapkan terima kasih kepada:

1. Allah SWT yang senantiasa memberikan karunia sehingga skripsi ini dapat diselesikan dengan baik dan lancar.
2. Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
3. Bapak Sudarmawan, M.T selaku ketua jurusan S1 Informatika Universitas Amikom Yogyakarta.
4. Bapak Yoga Pristyanto, S.Kom, M.Eng yang telah memberikan opsi pilihan tema penelitian skripsi.
5. Bapak Arif Akbarul Huda, S.Si, M.Eng yang senantiasa membimbing dalam mengerjakan skripsi ini hingga selesai.
6. Bapak Dony Ariyus, M.Kom, dan Theopilus Bayu Sasongko, S.Kom, M.Eng yang telah memberikan saran-saran.
7. Bapak dan ibu dosen yang telah memberikan ilmu dan motivasi selama perkuliahan.
8. Para penulis sumber bacaan, jurnal, dan makalah yang penulis jadikan referensi dalam penulisan laporan skripsi ini.
9. Ibu Sulis wali kelas di MAN 2 Yogyakarta, yang senantiasa memberikan doa dan nasihat yang sangat berarti.



10. Teman-teman IPA 1 2017, yang selalu saling support satu sama lainnya.
11. Sahabat-sahabatku ketika di MAN 2 Yogyakarta hingga sekarang, Basya Ali, Dian Afnita, dan Izah Fadhila yang selalu memberikan semangat dan kebahagiaan.
12. Sahabat-sahabatku ketika kuliah Toti, Rangga, Aang, Fery, Imeh, Gica, Farida, dan semua anak IF12 2017 yang tidak bisa disebutkan semuanya.
13. Laptop kentang yang senantiasa menemani dari awal kuliah hingga lulus.
14. JKT48 yang telah memberikan aura semangat yang positif sehingga termotivasi untuk menyelesaikan skripsi ini dengan cepat serta menjalani hidup penuh dengan semangat melalui karya-karyanya.

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.

Yogyakarta, 23 September 2020

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>II</b>
<b>PERSETUJUAN.....</b>	<b>III</b>
<b>PENGESAHAN.....</b>	<b>IV</b>
<b>PERNYATAAN.....</b>	<b>V</b>
<b>MOTTO .....</b>	<b>VI</b>
<b>PERSEMBAHAN.....</b>	<b>VII</b>
<b>KATA PENGANTAR.....</b>	<b>VIII</b>
<b>DAFTAR ISI.....</b>	<b>X</b>
<b>DAFTAR TABEL.....</b>	<b>XII</b>
<b>DAFTAR GAMBAR.....</b>	<b>XIII</b>
<b>INTISARI .....</b>	<b>XIV</b>
<b>ABSTRACT .....</b>	<b>XV</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 LATAR BELAKANG .....	1
1.2 RUMUSAN MASALAH.....	5
1.3 BATASAN MASALAH.....	6
1.4 MAKSUD DAN TUJUAN PENELITIAN.....	6
1.5 MANFAAT PENELITIAN .....	6
1.6 METODE PENELITIAN.....	7
1.7 SISTEMATIKA PENULISAN.....	8
<b>BAB II LANDASAN TEORI.....</b>	<b>9</b>
2.1 KAJIAN PUSTAKA .....	9
2.2 OPTIMASI.....	12
2.3 <i>WEBSITE PHISHING</i> (SITUS PHISHING).....	12
2.4 ALGORITMA.....	14
2.5 KLASIFIKASI .....	15
2.6 FITUR DETEKSI.....	18
2.7 DECISION TREE.....	19
2.8 ALGORITMA CART .....	19
2.9 BAGGING .....	20

2.10	BAHASA PEMROGRAMAN PYTHON.....	21
2.11	FLASK.....	22
<b>BAB III METODE PENELITIAN .....</b>		<b>25</b>
3.1	ALAT DAN BAHAN PENELITIAN .....	25
3.1.1	Alat.....	25
3.1.1.1	Perangkat Keras.....	25
3.1.1.2	Perangkat Lunak.....	25
3.1.2	Bahan Penelitian.....	26
3.2	ALUR PENELITIAN .....	26
3.2.1	Studi Literatur .....	27
3.2.2	Pengumpulan Data .....	27
3.2.3	<i>Preprocessing</i> Data .....	27
3.2.4	Uji Coba dan Perbandingan Metode.....	28
3.2.5	Perbaikan Metode dan Evaluasi .....	28
3.2.6	Implementasi Sistem .....	29
3.2.7	Hasil dan Kesimpulan .....	29
3.2.8	Diagram Alur Penelitian.....	29
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>		<b>31</b>
4.1	HASIL PENGUJIAN.....	31
4.1.1	Hasil Pengujian <i>Decision Tree</i> (CART).....	31
4.1.2	Hasil Pengujian <i>Decision Tree</i> (CART) <i>Bagging</i> .....	36
4.2	ANALISIS HASIL PENGUJIAN.....	40
4.2.1	<i>Accuracy</i> (Akurasi).....	40
4.2.2	<i>Precision, Recall, dan F-Measure Phishing</i> .....	41
4.2.3	<i>Precision, Recall, dan F-Measure Non-Phishing</i> .....	42
4.3	SISTEM DETEKSI <i>WEBSITE PHISHING</i> .....	44
4.3.1	Uji Coba Sistem Deteksi <i>Website Phishing</i> .....	46
4.3.2	Hasil Uji Coba Sistem Deteksi <i>Website Phishing</i> .....	47
<b>BAB V PENUTUP.....</b>		<b>49</b>
5.1	KESIMPULAN .....	49
5.2	SARAN .....	50
<b>DAFTAR PUSTAKA .....</b>		<b>51</b>

## DAFTAR TABEL

Tabel 2.1 Hasil Perbandingan Algoritma C4.5 Dan Genetika.....	9
Tabel 2.2 Hasil Perbandingan Kinerja Algoritma SMO, Naïve Bayes, Bagging, Dan Multilayer Perceptron.....	10
Tabel 2.3 Persamaan Dan Perbedaan Penelitian Sebelumnya.....	11
Tabel 2.4 Atribut Dan Label Klasifikasi <i>Website Phishing</i> .....	16
Tabel 3.1 <i>Confusion Matrix</i> .....	28
Tabel 4.1 <i>Hyperparameter Value Train Test Split</i> Algoritma <i>Decision Tree</i> (CART) .....	32
Tabel 4.2 <i>Hyperparameter</i> Algoritma <i>Decision Tree</i> (CART) .....	32
Tabel 4.3 <i>Confusion Matrix</i> Algoritma <i>Decision Tree</i> (CART).....	33
Tabel 4.4 Kinerja Algoritma <i>Decision Tree</i> (CART).....	34
Tabel 4.5 <i>Hyperparameter Value Train Test Split</i> Algoritma <i>Decision Tree</i> (CART) <i>Bagging</i> .....	36
Tabel 4.6 <i>Hyperparameter</i> Algoritma <i>Decision Tree</i> (CART) <i>Bagging</i> .....	36
Tabel 4.7 <i>Confusion Matrix</i> Algoritma <i>Decision Tree</i> (CART) <i>Bagging</i> .....	37
Tabel 4.8 Kinerja Algoritma <i>Decision Tree</i> (CART) <i>Bagging</i> .....	38
Tabel 4.9 Hasil Kinerja Akurasi Algoritma <i>Decision Tree</i> (CART) Dan <i>Decision Tree</i> (CART) <i>Bagging</i> .....	40
Tabel 4.10 <i>Precision, Recall, Dan F-Measure Phishing</i> .....	41
Tabel 4.11 <i>Precision, Recall, Dan F-Measure Non-Phishing</i> .....	42
Tabel 4.12 Data Sampel URL <i>Website Phishing Dan Non-Phishing</i> .....	46
Tabel 4.13 Hasil Deteksi Data Sampel URL <i>Website Phishing Dan Non-Phishing</i> .....	48

## DAFTAR GAMBAR

Gambar 1.1 Perbandingan <i>Website Phishing</i> Dengan <i>Non-Phishing</i> .....	3
Gambar 1.2 Grafik <i>Phishing Activity Trend</i> Kuartal Pertama 2020.....	4
Gambar 2.1 Perbandingan Akurasi Algoritma <i>Decision Tree</i> , <i>Naïve Bayes</i> , Dan <i>Svm</i> .....	11
Gambar 2.2 Hubungan Masalah, Algoritma, Dan Solusi .....	14
Gambar 2.3 Proses Klasifikasi .....	18
Gambar 2.4 Alur Diagram Metode <i>Bagging</i> .....	21
Gambar 3.1 Alur Penelitian.....	30
Gambar 4.1 Perbandingan Akurasi Algoritma <i>Decision Tree (CART)</i> Dan <i>Decision Tree (CART) Bagging</i> .....	41
Gambar 4.2 Perbandingan Nilai <i>Precision</i> , <i>Recall</i> , Dan <i>F-Measure Phishing</i> .....	42
Gambar 4.3 Perbandingan Nilai <i>Precision</i> , <i>Recall</i> , Dan <i>F-Measure Non-Phishing</i> .....	43
Gambar 4.4 <i>Interface</i> Sistem Deteksi <i>Website Phishing</i> .....	44
Gambar 4.5 Hasil Deteksi Jika Prediksi Sebagai <i>Website Non-Phishing</i> .....	45
Gambar 4.6 Hasil Deteksi Jika Prediksi Sebagai <i>Website Phishing</i> .....	45
Gambar 4.7 Alur Diagram Proses Sistem Deteksi <i>Website Phishing</i> .....	46

## INTISARI

Perkembangan Ilmu Pengetahuan dan Teknologi (IPTEK), terutama pada Teknologi Informasi memberi kemudahan manusia dalam menunjang aktivitasnya, disisi lain berbagai kemudahan menjadi celah keamanan terhadap pemanfaatan oleh penjahat dunia maya. Salah satunya kegiatan yaitu mencuri data dari pengguna internet melalui sebuah *website* palsu yang dibangun seperti asli atau disebut juga *website phishing*.

Pada penelitian ini, untuk mengatasi maraknya *website phishing* yang ada pada dunia maya maka dibutuhkan suatu klasifikasi untuk memprediksi *website* yang terindikasi *phishing* dengan menggunakan algoritma klasifikasi *decision tree* (CART). Untuk meningkatkan kinerja dari algoritma klasifikasi *decision tree* (CART) agar dihasilkan akurasi yang lebih baik dan optimal, maka diperlukan optimasi yaitu dengan menggunakan salah satu algoritma penggabungan dari *machine learning* yaitu metode *bagging*.

Berdasarkan hasil uji coba, penerapan algoritma klasifikasi *decision tree* (CART) dalam klasifikasi *website phishing* dihasilkan akurasi sebesar 96,61%, setelah dilakukan penggabungan dengan metode *bagging* akurasi yang dihasilkan naik sebesar 1,13% sehingga menjadi 97,74%. Dari uji coba ini terbukti bahwa optimasi dengan melakukan penggabungan antara algoritma *decision tree* (CART) dan metode *bagging* dapat meningkatkan akurasi dalam prediksi *website phishing*.

**Kata Kunci:** *website phishing*, klasifikasi, prediksi, algoritma klasifikasi *decision tree* (CART), metode *bagging*.

## **ABSTRACT**

*The development of Science and Technology, especially in Information Technology, makes it easy for humans to support their activities, on the other hand various facilities become security holes against exploitation by cyber criminals. One of the activities is stealing data from internet users through a fake website that is built like a real one or also called a phishing website.*

*In this study, to overcome the prevalence of phishing websites in cyberspace, a classification is needed to predict phishing-indicated websites using decision tree (CART) classification algorithms. To improve the performance of the decision tree classification algorithm (CART) in order to produce better and optimal accuracy, optimization is needed by using one of the combining algorithms of machine learning is the bagging method.*

*Based on the test results, the application of the decision tree (CART) classification algorithm in the classification of phishing websites resulted in an accuracy of 96.61%, after merging with the bagging method the resulting accuracy increased by 1.13% to become 97.74%. From this trial, it is evident that optimization by combining the decision tree (CART) algorithm and the bagging method can improve accuracy in predicting phishing websites.*

**Keyword:** *phishing website, classification, predict, decision tree (CART) classification algorithms, bagging method.*