



**DETEKSI SITUS PHISING MENGGUNAKAN *RANDOM FOREST* DENGAN
TREE PARZEN ESTIMATOR *HYPER-PARAMETER* TUNNING**

SKRIPSI

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

MUHAMMAD ALDO REVALDY

NIM 1711016210012

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JUNI 2023



**DETEKSI SITUS PHISING MENGGUNAKAN *RANDOM FOREST* DENGAN
TREE PARZEN ESTIMATOR *HYPER-PARAMETER* TUNNING**

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

MUHAMMAD ALDO REVALDY

NIM 1711016210012

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2023**

SKRIPSI

**DETEKSI SITUS PHISING MENGGUNAKAN RANDOM FOREST
DENGAN TREE PARZEN ESTIMATOR HYPER-PARAMETER TUNING**

Oleh :

MUHAMMAD ALDO REVALDY

1711016210012

Telah dipertahankan di depan Dosen Penguji pada tanggal 23 Juni 2023,


Susunan Dosen Penguji :

Pembimbing I



Muhammad Itqan Mazdadi,
S.Kom., M.Kom
NIP. 199006122019031013

Dosen Penguji I



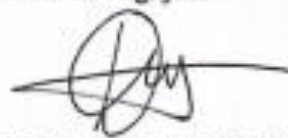
Dodon Turianto Nugrahadi,
S.Kom., M.Eng
NIP. 198001122009121002

Pembimbing II



Rudy Hertono, S.Kom., M.Kom
NIP. 198809252022031003

Dosen Penguji II



Rahmat Ramadhani, S.Kom., M.Sc.
NIP. 19920330201901110001

Banjarbaru, 23 Juni 2023

Koordinator Program Studi Ilmu Komputer



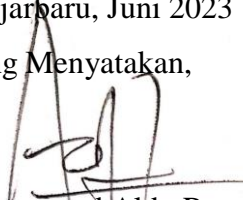
Irwan Budiman, S.T., M.Kom
NIP. 197703252008121001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, Juni 2023

Yang Menyatakan,



Muhammad Aldo Revaldy

NIM. 1711016210012

ABSTRAK

DETEKSI SITUS PHISING MENGGUNAKAN *RANDOM FOREST* DENGAN *TREE PARZEN ESTIMATOR HYPER-PARAMETER TUNNING*

(Oleh: Muhammad Aldo Revaldy; Pembimbing: Muhammad Itqan Mazdadi, S.Kom., M.Kom dan Rudy Herteno, S.Kom., M.Kom ; 2022; 62 halaman)

Phising adalah salah satu bentuk *Cyber Crime* yang merugikan orang lain dan termasuk perbuatan yang melawan hukum. Terdapat beberapa pendekatan untuk memberantas kejahatan *phising*, salah satunya dengan melakukan klasifikasi terhadap situs *phising* menggunakan metode *machine learning*. Dataset yang digunakan yaitu dataset situs *phising* dari *UCI Repository* dengan 11055 data dan 30 fitur kategorial. Metode pengklasifikasi yang digunakan yaitu *Random Forest*. *Random Forest* baik untuk melakukan klasifikasi terhadap data dengan fitur kategorikal, namun kinerja algoritma ini masih bisa ditingkatkan. Untuk mengatasi permasalahan tersebut peneliti menggunakan solusi *hyper-parameter tuning*. *Random Forest* memiliki beberapa *hyper-parameter* yang dapat dikonfigurasi untuk meningkatkan kinerja dari model. Masalah mengidentifikasi nilai yang baik untuk *hyper-parameter* disebut *hyper-parameter tuning*. Metode *hyper-parameter tuning* yang digunakan yaitu *Tree Parzen Estimator* yang kemudian divalidasi menggunakan *5-Fold Cross Validation* sebanyak 150 iterasi. *Hyper-parameter Random Forest* yang dikonfigurasi antara lain *n_estimators*, *max_depth*, *min_samples_leaf* dan *max_features*. Pengujian pada *Random Forest* tanpa *hyper-parameter tuning* memperoleh akurasi sebesar 97,06%. Untuk pengujian *Random Forest* dengan *hyper-parameter tuning* memperoleh akurasi sebesar 97,2%. *Hyper-parameter tuning* dengan *Tree Parzen Estimator* pada *Random Forest* untuk klasifikasi situs *phising* memberikan peningkatan kinerja model pada akurasi sekitar 0,14%.

Kata kunci: *Phising, Random Forest, Hyper-parameter Tuning, Tree Parzen Estimator*

ABSTRACT

PHISHING SITE DETECTION USING RANDOM FOREST WITH TREE PARZEN ESTIMATOR HYPER-PARAMETER TUNNING

(By: Muhammad Aldo Revaldy; Supervisor: Muhammad Itqan Mazdadi, S.Kom., M.Kom dan Rudy Herteno, S.Kom., M.Kom; 2022; 62 pages)

*Phishing is a form of cybercrime that harms others and is considered illegal. There are several approaches to combatting phishing, one of which is to classify phishing sites using machine learning methods. The dataset used is the phishing site dataset from the UCI Repository, which consists of 11,055 data points and 30 categorical features. The classification method used is Random Forest, which is suitable for classifying data with categorical features, but its algorithm's performance can still be improved. To address this issue, the researchers used a hyperparameter tuning solution. Random Forest has several hyperparameters that can be configured to improve the model's performance. The problem of identifying good values for hyperparameters is called hyperparameter tuning. The hyperparameter tuning method used is Tree Parzen Estimator, which is then validated using 5-Fold Cross-Validation for 150 iterations. The hyperparameters configured for Random Forest include *n_estimators*, *max_depth*, *min_samples_leaf*, and *max_features*. Testing Random Forest without hyperparameter tuning obtained an accuracy of 97.06%. Testing Random Forest with hyperparameter tuning obtained an accuracy of 97.2%. Hyperparameter tuning with Tree Parzen Estimator on Random Forest for phishing site classification improved the model's performance by around 0.14% in accuracy.*

Keywords: *Phishing, Random Forest, Hyper-parameter Tuning, Tree Parzen Estimator.*

PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh.

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Deteksi Situs Phising Menggunakan *Random Forest* Dengan *Tree Parzen Estimator Hyper-Parameter Tuning*” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan shalawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga *yaumul qiyamah*.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

1. Keluarga terutama kepada Orang tua yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini.
2. Bapak Muhammad Itqan Mazdadi, S.Kom., M.Kom selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran penyelesaian skripsi ini.
3. Bapak Rudy Herteno, S.Kom., M.Kom selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman, S.T., M.Kom selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA UNLAM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman keluarga Ilmu Komputer angkatan 2017, terima kasih untuk canda, tawa, perjuangan yang sudah dilewati bersama, untuk semua kenangan manis yang telah terukir selama ini. Senang bisa menjadi salah satu bagian dari kehidupan kalian.

7. Wanita yang sangat saya sayangi Rahmah Hayati yang telah memberi support dan selalu ada selalu bisa seperti Tokopedia
8. Teman-teman anggota Seven Deadly Sins (七つの大罪) Said, Ryan, Haekal, Uzik, Lutfi, Barok yang telah bersedia membantu memecahkan kebingungan-kebingungan dan memberikan saran, dukungan, semangat, serta motivasi, selama proses penyelesaian skripsi.
9. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, Juni 2023

Muhammad Aldo Revaldy

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	Error! Bookmark not defined.
HALAMAN PERNYATAAN	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xii
BAB I PENDAHULUAN.....	Error! Bookmark not defined.
1.1 Latar Belakang.....	Error! Bookmark not defined.
1.2 Rumusan Masalah.....	Error! Bookmark not defined.
1.3 Batasan Masalah	Error! Bookmark not defined.
1.4 Tujuan.....	Error! Bookmark not defined.
1.5 Manfaat Penelitian.....	Error! Bookmark not defined.
BAB II TINJAUAN PUSTAKA.....	Error! Bookmark not defined.
2.1 Kajian Terdahulu	Error! Bookmark not defined.
2.2 Landasan Teori	Error! Bookmark not defined.
2.2.1 Phising.....	Error! Bookmark not defined.
2.2.2 Phishing Detection	Error! Bookmark not defined.
2.2.3 Decision Tree	Error! Bookmark not defined.
2.2.4 <i>Ensemble</i> Learning.....	Error! Bookmark not defined.
2.2.5 Random Forest	Error! Bookmark not defined.

2.2.6	<i>Hyper-Parameter</i>	Error! Bookmark not defined.
2.2.7	<i>Hyper-Parameter Tuning</i>	Error! Bookmark not defined.
2.2.8	<i>Tree Parzen Estimator (TPE)</i>	Error! Bookmark not defined.
2.2.9	<i>Crossvalidation</i>	Error! Bookmark not defined.
2.2.10	<i>Confusion Matrix</i>	Error! Bookmark not defined.
BAB III METODE PENELITIAN		Error! Bookmark not defined.
3.1	Bahan Penelitian	Error! Bookmark not defined.
3.2	dan Alat Penelitian.....	Error! Bookmark not defined.
3.3	Variabel Penelitian.....	Error! Bookmark not defined.
3.4	Prosedur Penelitian	Error! Bookmark not defined.
BAB IV HASIL DAN PEMBAHASAN		Error! Bookmark not defined.
4.1	Hasil.....	Error! Bookmark not defined.
4.1.1	Pengumpulan Dataset.....	Error! Bookmark not defined.
4.1.2	Modeling	Error! Bookmark not defined.
4.1.3	Evaluasi Hasil Kerja Klasifikasi	Error! Bookmark not defined.
4.2	Pembahasan	Error! Bookmark not defined.
BAB V PENUTUP		Error! Bookmark not defined.
5.1	Kesimpulan.....	Error! Bookmark not defined.
5.2	Saran	Error! Bookmark not defined.
DAFTAR PUSTAKA		Error! Bookmark not defined.
RIWAYAT PENULIS		Error! Bookmark not defined.

DAFTAR TABEL

Tabel	Halaman
Tabel 1 Keaslian Penelitian.....	Error! Bookmark not defined.
Tabel 2 Perancangan Penelitian	Error! Bookmark not defined.
Tabel 3 Parameter Random Forest	Error! Bookmark not defined.
Tabel 4 <i>Confusion Matrix Binary Classification</i>	Error! Bookmark not defined.
Tabel 5 Fitur pada Phising Website Data Set.....	Error! Bookmark not defined.
Tabel 6 Contoh dataset Phising Website Data Set	Error! Bookmark not defined.
Tabel 7 Pembagian <i>data training</i> dan <i>data testing</i>	Error! Bookmark not defined.
Tabel 8 <i>Hyper-parameter Random Forest</i>	Error! Bookmark not defined.
Tabel 9 Domain pencarian hyper-parameter Random Forest	Error! Bookmark not defined.
Tabel 10. Kandidat <i>hyper-parameter tree parzen estimator</i>	Error! Bookmark not defined.
Tabel 11 <i>5-Fold Cross Validation</i> dari kandidat <i>hyper-parameter</i>	Error! Bookmark not defined.
Tabel 12 <i>Confusion matrix</i> klasifikasi <i>Random Forest</i> tanpa <i>hyper-parameter tuning</i>	Error! Bookmark not defined.
Tabel 13 Perhitungan akurasi klasifikasi <i>Random Forest</i> tanpa <i>hyper-parameter tuning</i>	Error! Bookmark not defined.
Tabel 14 <i>Confusion matrix</i> klasifikasi <i>Random Forest</i> dengan <i>hyper-parameter tuning</i>	Error! Bookmark not defined.
Tabel 15 Perhitungan akurasi klasifikasi <i>Random Forest</i> dengan <i>hyper-parameter tuning</i>	Error! Bookmark not defined.

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1 Anatomi <i>URL</i>	Error! Bookmark not defined.
Gambar 2 Model hierarki prediksi situs <i>phising</i>	Error! Bookmark not defined.
Gambar 3 Alur <i>Machine Learning</i>	Error! Bookmark not defined.
Gambar 4 Struktur pohon pada decision tree	Error! Bookmark not defined.
Gambar 5. Alur Kerja Random Forest	Error! Bookmark not defined.
Gambar 6 Ilustrasi <i>Split Validation</i>	Error! Bookmark not defined.
Gambar 7. Penentuan grup TPE.....	Error! Bookmark not defined.
Gambar 8. Alur cara kerja Tree Parzen Estimator	Error! Bookmark not defined.
Gambar 9. Contoh Simulasi <i>Crossvalidation</i>	Error! Bookmark not defined.
Gambar 10. Alur Penelitian.....	Error! Bookmark not defined.
Gambar 11 Distribusi kelas pada <i>Phising</i> Website Data Set	Error! Bookmark not defined.
Gambar 12 Perbandingan persentase kelas <i>phising</i> dengan kelas non <i>phising</i> ...	Error! Bookmark not defined.
Gambar 13 Visualisasi <i>Phising</i> Website Data Set	Error! Bookmark not defined.
Gambar 14 Heatmap Korelasi Tiap Fitur <i>Phising</i> Website Data Set.....	Error! Bookmark not defined.
Gambar 15 Ilustrasi pembagian dataset	Error! Bookmark not defined.
Gambar 16 Alur klasifikasi <i>Random Forest</i> tanpa <i>hyper-parameter tuning</i>	Error! Bookmark not defined.
Gambar 17 Root Node Pohon Keputusan <i>Random Forest</i>	Error! Bookmark not defined.
Gambar 18 Node 1.2 Pohon Keputusan <i>Random Forest</i>	Error! Bookmark not defined.
Gambar 19 Pohon Keputusan Pertama <i>Random Forest</i>	Error! Bookmark not defined.
Gambar 20 Pohon keputusan kedua <i>Random Forest</i> .	Error! Bookmark not defined.
Gambar 21 Pohon keputusan ketiga <i>Random Forest</i> .	Error! Bookmark not defined.

Gambar 22. Alur klasifikasi *Random Forest* dengan *hyper-parameter tuning* ..**Error! Bookmark not defined.**

Gambar 23 Pembagian data dengan *5-Fold Cross Validation***Error! Bookmark not defined.**

Gambar 24 Hasil dari *Tree Parzen Estimator* dengan 150 iterasi**Error! Bookmark not defined.**

Gambar 25 Perbandingan kinerja model..... **Error! Bookmark not defined.**

DAFTAR LAMPIRAN

Lampiran

Lampiran 1 Import dataset phishing dan pembagian dataset

Lampiran 2 Source code confusion matrix

Lampiran 3 Klasifikasi *Random Forest* tanpa *hyper-parameter tuning*

Lampiran 4 Sourcode klasifikasi *Random Forest* dengan *hyper-parameter tuning*

