

**KLASIFIKASI MALWARE PDF MENGGUNAKAN DEEP NEURAL  
NETWORK PADA LAYANAN GARUDA KEMENDIKBUD SEBAGAI  
AGREGATOR NASIONAL**

**SKRIPSI**

**Diajukan Untuk Melengkapi Salah Satu  
Syarat Memperoleh Gelar Sarjana**



**Oleh :**

**RISKY WAHYUNI**

**09011382025158**

**JURUSAN SISTEM KOMPUTER**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS SRIWIJAYA**

**2024**

LEMBAR PENGESAHAN

**KLASIFIKASI MALWARE PDF MENGGUNAKAN  
DEEP NEURAL NETWORK PADA LAYANAN GARUDA  
KEMENDIKBUD SEBAGAI AGREGATOR NASIONAL**

**Skripsi**  
**Program Studi Sistem Komputer**  
**Jenjang S1**

**Oleh:**  
**RISKY WAHYUNI**  
**09011382025158**

Palembang, *26* Agustus 2024

**Pembimbing I**



**Prof. Deris Stiawan, M.T., Ph.D.**  
**NIP. 197806172006041002**

**Pembimbing II**



**Nurul Afifah, M.Kom.**  
**NIP. 199211102023212049**

**Mengetahui,**  
**Ketua Jurusan Sistem Komputer**



**Dr. Ir. Sukemi, M.T.**  
**NIP. 196612032006041001**

**AUTHENTICATION PAGE**

**PDF MALWARE CLASSIFICATION USING DEEP  
NEURAL NETWORK ON GARUDA KEMENDIKBUD  
SERVICE AS A NATIONAL AGGREGATOR**

**THESIS**

**Dept. of Computer System**

**Bachelor's Degree**

**By:**

**RISKY WAHYUNI**

**09011382025158**

**Palembang, 26 August 2024**

**Supervisor**



**Prof. Deris Stawan, M.T., Ph.D.**  
**NIP. 197806172006041002**

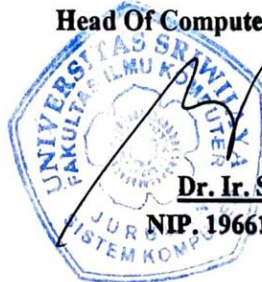
**Co-Supervisor**



**Nurul Afifah, M.Kom.**  
**NIP. 199211102023212049**

**Acknowledge, 26/8/24**

**Head Of Computer Systems Departement**



**Dr. Ir. Sukemi, M.T.**

**NIP. 196612032006041001**

## LEMBAR PERSETUJUAN

Telah di uji dan lulus pada:

Hari : Selasa

Tanggal : 24 Juli 2024

Tim Penguji :

1. Ketua : Aditya Putra Perdana Prasetyo, M.T. 

2. Sekretaris : Iman Saladin B. Azhar, M.MSL 

3. Pembimbing I : Prof. Deris Stiawan, M.T., Ph.D. 

4. Pembimbing II : Nurul Affah, M.Kom.

5. Penguji : Huda Ubaya, S.T., M.T.

Mengetahui, <sup>26/8/24</sup>  
Ketua Jurusan Sistem Komputer



**Dr. Ir. Sukemi, M.T.**

**NIP. 196612032006041001**

## Halaman Pernyataan

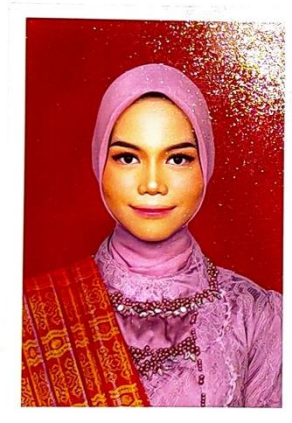
Yang Bertanda Tangan Dibawah Ini:

Nama : Risky Wahyuni  
Nim : 09011382025158  
Judul : Klasifikasi Malware PDF Menggunakan Deep  
Network Pada Layanan Garuda Kemendikbud Sebagai  
Agregator Nasional.

**Hasil Pengecekan Software *Thenticate/Turnitin*: 2%**

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari universitas sriwijaya.

Demikian, pernyataan ini saya buat dalam keadaan sadar dan tanpa paksaan dari siapapun.



Palembang, 23 Agustus 2024



Risky Wahyuni  
Nim. 09011382025158

## KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh dengan Rahmat dan berkah Allah SWT beserta kedamaiannya.

Puji syukur kehadiran Allah SWT, penulis memohon kepada Allah SWT agar dapat menyelesaikan proposal tugas akhir ini yang berjudul **“Klasifikasi Malware PDF Menggunakan Deep Neural Network Pada Layanan GARUDA Kemendikbud Sebagai Agregator Nasional”**

Penulis ingin menyampaikan penghargaan atas pemikiran, saran, dan bantuan yang telah diberikan dalam penyusunan tugas akhir ini. Sehubungan dengan hal tersebut, penulis mengucapkan puji syukur kepada Allah SWT dan yang mulia :

1. Ayahanda Sunardi, yang selalu menjadi panutanku. Ibunda Murtika, malaikat pelindung yang tak pernah lelah mencintaiku. Kalian berdua telah memberikan segalanya untukku, dari kasih sayang yang tulus hingga dukungan yang tak pernah putus. Setiap tetes keringat dan air mata yang kalian curahkan untukku adalah bukti cinta yang tak terhingga. Semoga penulis bisa membahagiakan kalian dan membuat kalian bangga.
2. Bapak Prof Erwin, S.Si., M.Si., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Dr. Ir. Sukemi, M.T. selaku Ketua Jurusan Sistem Komputer Universitas Sriwijaya.
4. Bapak Prof. Deris Stiawan, M.T., Ph.D., IPU., ASEAN Eng. selaku Dosen Pembimbing I Tugas Akhir yang berkenan meluangkan waktu dalam membimbing, memberikan saran dan motivasi untuk penulis dalam menyelesaikan tugas akhir ini.
5. Ibu Nurul Afifah, M.Kom. selaku Dosen Pembimbing II Tugas Akhir yang berkenan meluangkan waktu dalam membimbing, memberikan saran dan motivasi untuk penulis dalam menyelesaikan tugas akhir ini.
6. Bapak Muhammad Ali Buchari, M.T. selaku Dosen Pembimbing

Akademik.

7. Mba Sari Admin Jurusan Sistem Komputer yang telah membantu dalam mengelola seluruh berkas administrasi.
8. Indra Wikcaksono, yang selalu mendukung penulis pada hari yang tidak mudah selama proses pengerjaan skripsi, terimakasih telah mendengarkan keluh kesah penulis dan berkontribusi banyak, Terimakasih telah menjadi bagian perjalanan penulis.
9. Teruntuk sahabat, Mayin, Anngun, dinoy, ridoks, depoy, Terimakasih telah menjadi sahabat yang baik, mari tetap bersama meski menjalani hidup masing masing, mari bertemu dan bercerita banyak hal, mari untuk selalu tetap saling menyangi dan selalu mendukung dalam segala hal yang baik.
10. Virginita, Indah, Krisna, Nisa, Cytia, dan Vanesa selaku sahabat yang selalu memberi semangat serta dukungannya. Sehat selaluuu guys ehehe
11. Risky Wahyuni, diri saya sendiri apresiasi sebesar besarnya karna telah bertanggung jawab meyelesaikan skripsi ini, terimakasih karna tidak menyerah dan menikmati setiap prosesnya.
12. Teman-teman seperjuangan SK Unggulan 20
13. Dan semua pihak yang tidak dapat disebutkan satu persatu.
14. Almamater.

penulis berharap semoga laporan tugas akhir ini dapat memberikan manfaat bagi kita semua, khususnya bagi mahasiswa fakultas ilmu komputer universitas sriwijaya, baik secara langsung maupun tidak langsung sebagai kontribusi dalam meningkatkan taraf hidup pengajaran dan penelitian.

Palembang, Agustus 2024  
Penulis,

Risky Wahyuni  
NIM. 09011382025158

# KLASIFIKASI MALWARE PDF MENGGUNAKAN DEEP NEURAL NETWORK PADA LAYANAN GARUDA KEMENDIKBUD SEBAGAI AGREGATOR NASIONAL

**Risky Wahyuni (09011382025158)**

*Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya*

*Email : [riskywahyuni2502@gmail.com](mailto:riskywahyuni2502@gmail.com)*

## ABSTRAK

Gerba Rujukan Digital (GARUDA) merupakan repositori yang mengelola berbagai karya ilmiah Indonesia, artikel yang tersedia pada repositori GARUDA diterbitkan dalam format PDF yang rentan terhadap serangan malware sehingga membuka peluang bagi *hacker* untuk melakukan kejahatan dengan mencuri data pribadi, merusak file, atau mengambil kendali sistem tanpa izin. Oleh karena itu, penelitian berfokus untuk membedakan file PDF yang aman dengan file PDF berbahaya (*Malware*) penulis akan mengekstrak metadata dari setiap file PDF menggunakan PDFiD, metadata tersebut sebagai ciri (fitur) untuk melakukan klasifikasi *multiclass* dengan menggunakan metode *Deep Neural Network*. Tujuannya adalah untuk mengklasifikasikan file PDF menjadi tiga kategori benign, *malware html*, *malware PDF*. Evaluasi model pada penelitian ini menggunakan *confusion matrix* dengan menghitung nilai akurasi, presisi, recall, dan f1-score, pada penelitian ini melakukan 5 kali skenario pengujian dengan tujuan untuk membandingkan hasil terbaik hyperparameter tuning dari 5 skenario, sehingga didapat hasil terbaik pada scenario 4 dengan nilai akurasi 99.45%, presisi 100%, recall 98.36% serta *f1-score* 99.17%.

**Kata kunci :** PDF Malware, PDFiD, *Multiclass*, *Deep Neural Network*.



# PDF MALWARE CLASSIFICATION USING DEEP NEURAL NETWORK ON GARUDA KEMENDIKBUD SERVICE AS A NATIONAL AGGREGATOR

**Risky Wahyuni (09011382025158)**

*Dept. of Computer System, Faculty of Computer Science, Universitas Sriwijaya*

*Email : [riskywahyuni2502@gmail.com](mailto:riskywahyuni2502@gmail.com)*

## ABSTRACT

*Gerba Rujukan Digital (GARUDA) is a repository that manages various Indonesian scientific works, articles available in the GARUDA repository are published in PDF format which is vulnerable to malware attacks thus opening up opportunities for hackers to commit crimes by stealing personal data, damaging files, or taking control of the system without permission. Therefore, the research focuses on distinguishing safe PDF files from malicious PDF files (Malware) The author will extract metadata from each PDF file using PDFiD, the metadata as a feature to perform multiclass classification using the Deep Neural Network method. The goal is to classify PDF files into three categories benign, html malware, PDF malware. Evaluation of the model in this study using confusion matrix by calculating the value of accuracy, precision, recall, and f1-score, in this study conducted 5 times the test scenario with the aim of comparing the best results of hyperparameter tuning from 5 scenarios, so that the best results were obtained in scenario 4 with an accuracy value of 99.45%, precision 100%, recall 98.36% and f1-score 99.17%.*

**Keywords: PDF Malware, PDFiD, Multiclass, Deep Neural Network.**

## DAFTAR ISI

	HALAMAN
<b>LEMBAR PENGESAHAN .....</b>	<b>ii</b>
<b>AUTHENTICATION PAGE.....</b>	<b>iii</b>
<b>LEMBAR PERSETUJUAN .....</b>	<b>iv</b>
<b>HALAMAN PERNYATAAN.....</b>	<b>v</b>
<b>KATA PENGANTAR.....</b>	<b>vi</b>
<b>ABSTRAK .....</b>	<b>viii</b>
<b>ABSTRACK.....</b>	<b>ix</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR GAMBAR .....</b>	<b>xiii</b>
<b>DAFTAR TABEL .....</b>	<b>xv</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah.....	3
1.3 Tujuan.....	3
1.4 Manfaat.....	3
1.5 Batasan Masalah.....	4
1.6 Metodologi Penelitian .....	4
1.7 Sistematika Penulisan.....	5
<b>BAB II TIJUAN PUSTAKA.....</b>	<b>7</b>
2.1 Pendahuluan .....	7
2.2 Penelitian terkanit.....	7
2.3 Landasan Teori .....	8
2.3.1 File PDF .....	8
2.3.2 Fitur PDF.....	9
2.3.3 PDF Malware .....	11
2.4 Ekstraksi Dataset .....	12
2.4.1 Virus total .....	12
2.4.2 PDFID .....	12

2.5 Dataset PDF Malware .....	13
2.6 Imbalance Dataset .....	13
2.7 Deep Neural Network.....	14
2.7.1 Arsitektur DNN .....	16
2.8 Evaluasi Performa Klasifikasi .....	17
<b>BAB III METODE PENELITIAN .....</b>	<b>19</b>
3.1 Pendahuluan .....	19
3.2 Kebutuhan Perangkat Lunak dan Perangkat Keras .....	19
3.2.1 Kebutuhan Perangkat Keras .....	19
3.2.2 Kebutuhan Perangkat Lunak .....	19
3.3 Kerangka Penelitian .....	20
3.4 Perancangan Sistem.....	22
3.5 Persiapan Dataset .....	22
3.6 Dataset.....	24
3.7 Data Understanding.....	25
3.8 Exploratory Data Analysis .....	25
3.9 Pre-processing .....	25
3.9.1 Feature Selection .....	25
3.9.2 Label Encoder .....	26
3.9.3 Normalisasi .....	26
3.9.4 Teknik Resampling.....	27
3.10 Klasifikasi Deep Neural Network .....	28
3.10.1 Parameter Pengujian.....	30
3.10.2 Program Pengujian Model.....	30
3.10.1 Arsitektur Deep Neural Network .....	31
<b>BAB IV ANALISA DAN HASIL .....</b>	<b>34</b>
4.1 Pendahuluan .....	34
4.2 Hasil Ekstraksi Dataset.....	34
4.3 Analisis Dastaset .....	35
a) File 696016.pdf.....	36
b) File 492775.pdf.....	36
c) File 496497.pdf.....	37

4.4 Data Understanding.....	40
4.5 Exploratory Data .....	41
4.6 Pre-Processing.....	42
4.6.1 Seleksi fitur .....	42
4.6.2 Label Encoder .....	43
4.6.3 Normalisasi .....	43
4.6.4 Teknik Resampling.....	43
4.7 Model Deep Neural Network .....	50
4.7.1 Pelatihan Model .....	50
4.8 Evaluasi Performa Klasifikasi .....	53
4.9 Evaluasi Model.....	56
4.9.1 Validasi Hasil Perhitungan Manual.....	57
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>62</b>
5.1 Kesimpulan.....	62
5.2 Saran.....	62
<b>DAFTAR PUSTAKA.....</b>	<b>63</b>

## DAFTAR GAMBAR

	HALAMAN
Gambar 2.1 Fitur PDF .....	8
Gambar 2. 2 Dataset PDF GARUDA .....	13
Gambar 2.3 Single Layer Neural Networks [18] .....	16
Gambar 2.4 Multiple Layer Neural Networks [18].....	17
Gambar 2. 5 Confusion Matrix .....	17
Gambar 3.1 Kerangka Penelitian .....	21
Gambar 3.2 Perancangan Sistem.....	22
Gambar 3.3 Alur Persiapan Dataset .....	23
Gambar 3.4 flowchart dataset .....	24
Gambar 3. 5 Flowchart Normalisasi .....	26
Gambar 3. 6 Teknik Resampling RUS dan ROS.....	28
Gambar 3.7 Model Deep Neural Network.....	29
Gambar 3. 8 Pseudocode Program Deep Neural Network 7 Hidden Layer.....	31
Gambar 3. 9 Arsitektur Deep Neural Network 7 Hidden Layer .....	32
Gambar 4.1 Tampilan PDF Normal .....	35
Gambar 4.2 Tampilan Dataset Malware.....	35
Gambar 4.3 Pdf Malware .....	36
Gambar 4.4 Malware HTML .....	37
Gambar 4.5 Pdf Benign.....	38
Gambar 4. 6 Ekstraksi Fitur Data Menggunakan PDFID .....	38
Gambar 4.7 pdf-Parser .....	38
Gambar 4.8 Fitur Dataset .....	40
Gambar 4.9 Jumlah Fitur Dengan Data Kosong .....	40
Gambar 4.10 Tipe Data Setiap Fitur .....	41
Gambar 4.11 Histogram Fitur Data.....	42
Gambar 4.12 hasil seleksi fitur.....	42
Gambar 4.13 Hasil Label Encoder .....	43
Gambar 4.14 Hasil Normalisasi .....	43
Gambar 4.15 Data Imbalance.....	44
Gambar 4. 16 Graph Accuracy dan Loss.....	45

Gambar 4. 17 Confusion Matrix Non Resampling .....	45
Gambar 4. 18 Data Random Understanding .....	46
Gambar 4. 19 Accuracy Kurve dan Loss .....	47
Gambar 4. 20 Counfusion Matrix .....	47
Gambar 4. 21 diagram pie data imbalance.....	48
Gambar 4. 22 Data Setelah Random Oversampling .....	48
Gambar 4.23 Data Balance .....	49
Gambar 4.24 Training Model Deep Neural Network.....	50
Gambar 4.25 Accuracy Curves dan Loss Best Performa Tiap Layer.....	52
Gambar 4.26 Hasil Rata Rata Akurasi .....	56
Gambar 4.27 Confusion Matrix .....	60

## DAFTAR TABEL

	<b>HALAMAN</b>
Tabel 2.1 Daftar Penelitian Terkait .....	7
Tabel 2. 2 Fitur PDF .....	9
Tabel 3.1 Kebutuhan Perangkat Lunak .....	19
Tabel 3.2 Kebutuhan Perangkat Keras .....	20
Tabel 3.3 hyparameter pengujian .....	30
Tabel 4.1 Tampilan Ekstraksi Fitur dalam Bentuk CSV .....	39
Tabel 4.2 Data Setelah Random Undesampling .....	46
Tabel 4.3 Distrubusi Kelas Setelah Random Oversampling .....	49
Tabel 4.4 Evaluasi Performa Model 4 Layer, 50 Epoch .....	54
Tabel 4.5 Evaluasi Performa Model 5 Layer, 100 Epoch .....	54
Tabel 4.6 Evaluasi Performa 6 Hidden Layer Epoch 150 .....	55
Tabel 4.7 Evaluasi Performa Model 7 Layer 200 Epoch .....	55
Tabel 4.8 Evaluasi Performa Model 8 Layer 250 Epoch .....	56

# BAB I PENDAHALUAN

## 1.1 Latar Belakang

Gerba Rujukan digital atau dikenal dengan GARUDA merupakan repositori yang menjadi sumber kumpulan informasi beragam seperti karya tulis ilmiah, tugas akhir mahasiswa, rangkuman studi, matematika, ilmu komputer dan berbagai materi lainnya, seluruh karya yang dikumpulkan direpositori GARUDA merupakan hasil karya anak bangsa Indonesia yang dikelola oleh kementerian ristek dan teknologi (KEMENDIKBUD DIKTI). GARUDA menyediakan lebih dari sajuta publikasi dan jurnal, ada lebih dari 40 bidang yang berbeda mulai dari astronomi, ilmu sosial, hingga Pendidikan [1]. Namun sama seperti platform umumnya, GARUDA juga memiliki risiko keamanan terutama terkait serangan siber. Artikel yang tersedia pada repositori ini diterbitkan dalam format PDF yang rentan terhadap serangan malware sehingga membuka peluang bagi *hacker* untuk melakukan kejahatan dengan mencuri data pribadi, merusak file, atau mengambil kendali sistem tanpa izin. Malware ini bisa menyebar melalui email, situs web, atau perangkat penyimpanan portable [2], Hacker dapat memodifikasi file PDF dengan menyuntikkan muatan berbahaya melalui objek lain dan kompleksitas format PDF untuk menyembunyikan konten berbahaya. Contohnya, serangan berbasis JavaScript menyuntikkan kode di berbagai lokasi dalam file untuk menipu pendeteksi, hacker juga menggunakan teknik pengelakan fitur untuk membuat varian malware PDF yang tidak terdeteksi oleh pengklasifikasi yang ada [3].

Dengan demikian penggunaan Deep Neural Network merupakan langkah penting dalam meningkatkan keamanan sistem terhadap ancaman malware, Deep Neural Network sangat akurat dalam menentukan kebenaran dalam malware dengan klasifikasi, terutama saat data tidak seimbang deep neural network mampu memberikan kinerja yang unggul dalam mempelajari fitur secara otomatis untuk menyelesaikan masalah kompleks seperti klasifikasi gambar, teks, dan audio dengan menghasilkan akurasi yang tinggi. Dalam beberapa tahun terakhir deep neural network telah menjadi salah satu



algoritma paling kuat dan efisien diberbagai bidang termasuk dalam mengklasifikasi malware [4].

Penelitian [5] menggunakan dataset yang terdiri dari 7414 file, dengan 3707 file malware dan 3707 file jinak. Dataset tersebut dilabeli sebagai X dan Y. Penelitian ini mengembangkan model Deep Neural Network (DNN) yang menunjukkan performa terbaik dibandingkan dengan teknik klasifikasi lainnya, serta mampu mengatasi masalah ketidakseimbangan dan overfitting data. Hasil penelitian menunjukkan bahwa model DNN dapat mengklasifikasikan perangkat lunak berbahaya dengan tingkat presisi, akurasi, dan recall yang tinggi, dengan akurasi mencapai 99,42%.

Penelitian [6] menggunakan 1,1 juta sampel malware windows dari koleksi EMBER untuk mengembangkan algoritme deteksi berbasis biner dan tanda tangan dengan deep neural network dan hasilnya menunjukkan akurasi pengujian sebesar 98,53%, mengatasi ketidak seimbangan data antara dataset kecil dan besar. Kesimpulannya, pendekatan deep learning ini dapat menjadi mesin algoritma utama untuk melindungi perangkat digital dari malware dan meningkatkan keamanan siber secara keseluruhan.

Penelitian [7] dataset 10.070 sampel, terdiri dari 3.961 sampel jinak dan 6.109 sampel berbahaya. sampel berbahaya dikumpulkan dari VirusTotal. Kumpulan data tersebut dipisahkan menjadi kumpulan data pelatihan A dan pengujian B. Studi ini bertujuan untuk mengklasifikasikan sampel malware sebagai benign atau malicious menggunakan metode deep learning, terutama CNN-BiLSTM, dengan fokus pada keberhasilan deteksi malware, ketahanan terhadap perubahan konsep, obfuscation packing, dan serangan adversarial. Hasil eksperimen menunjukkan bahwa metode yang diusulkan efektif dalam mendeteksi malware, terutama dalam menghadapi sampel malware yang lebih baru. Model-model yang diuji, seperti CNN-LSTM dan CNN-BiLSTM, menunjukkan kinerja yang baik dalam deteksi malware, dengan CNN-BiLSTM memiliki akurasi tertinggi dalam validasi silang. Metode yang diusulkan juga terbukti lebih tahan terhadap interferensi packing dan serangan adversarial dibandingkan dengan metode berbasis panggilan API Akurasi

tertinggi yang dicapai adalah 95.7% oleh model CNN-BiLSTM pada panjang input 2 juta.

Penelitian ini bertujuan untuk mengidentifikasi malware dalam file PDF menggunakan karakteristik khusus dari file PDF GARUDA. karakteristik pada PDF ini dieksekusi menggunakan *pdfid.py* untuk mengetahui dataframennya dan digunakan sebagai dataset untuk melatih model *Deep Neural Network* kemudian dilakukan klasifikasi dan analisis terhadap file PDF yang terindikasi mengandung malware. Sehingga pada Tugas akhir ini penulis akan melakukan penelitian dengan judul “*Klasifikasi Malware PDF Menggunakan Deep Neural Network Pada Layanan GARUDA Kemendikbud Sebagai Agregator Nasional*”

## **1.2 Perumusan Masalah**

1. Bagaimana proses ekstraksi file dataset PDF GARUDA sebelum di proses menjadi dataset?
2. Bagaimana pengaruh *Random Oversampling* sebagai metode resampling untuk menangani masalah data imbalance pdf malware GARUDA?
3. Bagaimana model *Deep Neural Networks* digunakan untuk mengklasifikasi PDF malware GARUDA?

## **1.3 Tujuan**

Adapun tujuan pada penulisan penelitian ini adalah :

1. Mengekstraksi dataset file PDF Malware GARUDA Menggunakan teknik analisis malware yaitu analisis statis.
2. Menerapkan teknik *Random Oversampling* pada dataset yang tidak seimbang untuk proses klasifikasi malware PDF.
3. Melakukan evaluasi performa deep neural network dalam klasifikasi serangan file PDF Malware.

## **1.4 Manfaat**

Adapun manfaat pada penulisan penelitian ini adalah :

1. Dataset GARUDA memberikan bahan uji beragam untuk mengembangkan dan mengevaluasi algoritma analisis malware berbasis analisis statis.

2. Memberikan Solusi untuk mengatasi ketidakseimbangan data pada dataset GARUDA, sehingga klasifikasi terhadap serangan file PDF menjadi lebih akurat.
3. Mengevaluasi seberapa baik performa model deep neural networks dalam mengklasifikasi file PDF GARUDA dengan memberikan akurasi yang baik.

### **1.5 Batasan Masalah**

Adapun manfaat pada penulisan penelitian ini adalah :

1. Berfokus pada Teknik ekstraksi yang khusus diterapkan pada dataset file pdf garuda.
2. Khusus mengimplementasi random oversampling pada satu model klasifikasi yaitu deep neural network.
3. Tidak melibatkan penggunaan metrik lainnya, evaluasi model deep neural network hanya menggunakan matrik evaluasi umum accuracy, precision, recall dan f1-score.
4. Tidak ada pembahasan mengenai upaya untuk mencegah serangan malware pdf.

### **1.6 Metodologi Penelitian**

Dalam penelitian tugas akhir ini akan dilakukan serangkaian tahapan dengan menerapkan metodologi sebagai berikut :

1. Pada tahap ini, penelitian dimulai dengan mengkaji literatur yang relevan sesuai dengan penelitian sebelumnya seperti artikel, makalah, jurnal terkait topik yang sama pada klasifikasi malware pdf menggunakan deep neural network.
2. Tahap perancangan sistem  
Tahap ini mencakup pemilihan perangkat keras dan perangkat lunak yang akan digunakan dalam penelitian.
3. Tahap pengujian  
Tahap pengujian dilakukan menggunakan parameter yang telah ditentukan sesuai dengan batasan masalah.
4. Tahap hasil dan Analisa

Hasil dari pengujian dianalisa untuk mengevaluasi kelebihan dan kekurangannya dengan demikian analisa ini dapat menghasilkan Kesimpulan dan saran sebagai referensi pada penelitian selanjutnya.

5. Tahap kesimpulan dan saran

Pada tahap akhir ini penulis Menyusun Kesimpulan dari seluruh kegiatan tugas akhir pada penelitian ini, penulis juga akan memberikan saran untuk penelitian yang akan datang.

### 1.7 Sistematika Penulisan

Berikut ini merupakan sistematika yang digunakan dalam penulisan tugas akhir, yang mendeskripsikan bab-bab yang terdapat dalam tugas akhir :

**BAB I**

**PENDAHULUAN**

Pada BAB I, akan membahas mengenai latar belakang, tujuan penelitian, manfaat penelitian, rumusan masalah, batasan masalah, metologi penelitian dan sistematika penulisan.

**BAB II**

**TINJAUAN PUSAKA**

Pada BAB II, mengkaji literatur yang relvan sesuai dengan penelitian sebelumnya seperti artikel, makalah, jurnal terkait topik yang sama pada klasifikasi malware pdf menggunakan deep neural network.

**BAB III**

**METOLOGI PENELITIAN**

Pada BAB III, akan menguraikan tahapan tahapan yang dilakukan dalam penelitian tugas akhir ini, serta menjelaskan rancangan sistem yang digunakan.

**BAB IV**

**HASIL DAN ANALISA**

Pada BAB IV, akan menyajikan hasil dari penelitian pada klasifikasi malware pdf menggunakan deep neural network, dengan memberikan Analisa dari algoritma yang digunakan.

**BAB V**

**PENDAHULUAN**

Pada BAB V, berisi Kesimpulan berdasarkan hasil penelitian yang telah dicapai, serta memberikan saran untuk pengembangan lebih lanjut dalam penelitian mendatang.

## DAFTAR PUSTAKA

- [1] R. Wahyudin, “Garba Rujukan Digital,” vol. 10, no. 1, pp. 62–63, 2010, [Online]. Available: <http://garuda.ristekbrin.go.id/documents/detail/480401>
- [2] C. Liu *et al.*, “A novel adversarial example detection method for malicious PDFs using multiple mutated classifiers,” *Forensic Sci. Int. Digit. Investig.*, vol. 38, p. 301124, 2021, doi: 10.1016/j.fsidi.2021.301124.
- [3] H. Bae, Y. Lee, Y. Kim, U. Hwang, S. Yoon, and Y. Paek, “Learn2Evade: Learning-Based Generative Model for Evading PDF Malware Classifiers,” *IEEE Trans. Artif. Intell.*, vol. 2, no. 4, pp. 299–313, 2021, doi: 10.1109/tai.2021.3103139.
- [4] E. Mbunge, B. Muchemwa, J. Batani, and N. Mbuyisa, “A review of deep learning models to detect malware in Android applications,” *Cyber Secur. Appl.*, vol. 1, no. April 2022, p. 100014, 2023, doi: 10.1016/j.csa.2023.100014.
- [5] N. Afifah and D. Stiawan, “The Implementation of Deep Neural Networks Algorithm for Malware Classification,” *Comput. Eng. Appl. J.*, vol. 8, no. 3, pp. 189–202, 2019, doi: 10.18495/comengapp.v8i3.294.
- [6] U. Divakarla, K. H. K. Reddy, and K. Chandrasekaran, “A Novel Approach towards Windows Malware Detection System Using Deep Neural Networks,” *Procedia Comput. Sci.*, vol. 215, no. 2022, pp. 148–157, 2022, doi: 10.1016/j.procs.2022.12.017.
- [7] W. Qiang, L. Yang, and H. Jin, “Efficient and Robust Malware Detection Based on Control Flow Traces Using Deep Neural Networks,” *Comput. Secur.*, vol. 122, p. 102871, 2022, doi: 10.1016/j.cose.2022.102871.
- [8] Y. Li, Y. Wang, Y. Wang, L. Ke, and Y. an Tan, “A feature-vector generative adversarial network for evading PDF malware classifiers,” *Inf. Sci. (Ny.)*, vol. 523, pp. 38–48, 2020, doi: 10.1016/j.ins.2020.02.075.
- [9] D. Gibert, J. Planes, C. Mateu, and Q. Le, “Fusing feature engineering and

- deep learning: A case study for malware classification,” *Expert Syst. Appl.*, vol. 207, no. April 2021, p. 117957, 2022, doi: 10.1016/j.eswa.2022.117957.
- [10] T. Tsafirir, A. Cohen, E. Nir, and N. Nissim, “Efficient feature extraction methodologies for unknown MP4-Malware detection using Machine learning algorithms,” *Expert Syst. Appl.*, vol. 219, no. May 2022, p. 119615, 2023, doi: 10.1016/j.eswa.2023.119615.
- [11] X. Ling *et al.*, “Adversarial attacks against Windows PE malware detection: A survey of the state-of-the-art,” *Comput. Secur.*, vol. 128, p. 103134, 2023, doi: 10.1016/j.cose.2023.103134.
- [12] A. Charim, S. Basuki, and D. R. Akbi, “Detect Malware in Portable Document Format Files (PDF) Using Support Vector Machine and Random Decision Forest,” *J. Online Inform.*, vol. 3, no. 2, p. 99, 2019, doi: 10.15575/join.v3i2.196.
- [13] A. Syukron and A. Subekti, “Penerapan Metode Random Over-Under Sampling dan Random Forest Untuk Klasifikasi Penilaian Kredit,” *J. Inform.*, vol. 5, no. 2, pp. 175–185, 2018, doi: 10.31311/ji.v5i2.4158.
- [14] E. M. B. Karbab, M. Debbabi, A. Derhab, and D. Mouheb, “MalDozer: Automatic framework for android malware detection using deep learning,” *DFRWS 2018 EU - Proc. 5th Annu. DFRWS Eur.*, vol. 24, pp. S48–S59, 2018, doi: 10.1016/j.diin.2018.01.007.
- [15] N. Verbiest, E. Ramentol, C. Cornelis, and F. Herrera, “Preprocessing noisy imbalanced datasets using SMOTE enhanced with fuzzy rough prototype selection,” *Appl. Soft Comput. J.*, vol. 22, pp. 511–517, 2014, doi: 10.1016/j.asoc.2014.05.023.
- [16] N. H. Latifah, A. Silvia, E. Prihatini, S. Nurmaini, and I. Yani, “Swarm Intelligent in Bio-Inspired Perspective: A Summary,” *Comput. Eng. Appl. J.*, vol. 7, no. 2, pp. 105–120, 2018, doi: 10.18495/comengapp.v7i2.255.
- [17] Wirogatama, “Arsitektur Neural Network,” *Archit. Neural Netw.*, pp. 181–

211, 2019.

- [18] “Knowledge is Power in Four Dimensions: Models to Forecast Future Paradigm,” *Knowledge is Power in Four Dimensions: Models to Forecast Future Paradigm*. 2022. doi: 10.1016/c2021-0-02583-9.
- [19] B. A. B. Ii and P. J. Komputer, “LANDASAN TEORI 1 . Local area network ( LAN ),” pp. 5–29, 2013, [Online]. Available: [https://repository.bsi.ac.id/index.php/unduh/item/2416/file\\_10-babII-landasan-teori.pdf](https://repository.bsi.ac.id/index.php/unduh/item/2416/file_10-babII-landasan-teori.pdf)
- [20] T. Wongvorachan, S. He, and O. Bulut, “A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining,” *Information*, vol. 14, p. 54, Jan. 2023, doi: 10.3390/info14010054.
- [21] M.-C. Popescu, V. Balas, L. Perescu-Popescu, and N. Mastorakis, “Multilayer perceptron and neural networks,” *WSEAS Trans. Circuits Syst.*, vol. 8, Jul. 2009.