DETEKSI ANOMALI FILE PDF MALWARE PADA LAYANAN AGREGATOR GARBA RUJUKAN DIGITAL (GARUDA) DENGAN ALGORITMA DECISION TREE

TUGAS AKHIR

OLEH :

NOVI YUNINGSIH

09011281823133

JURUSAN SISTEM KOMPUTER

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2022

DETEKSI ANOMALI FILE PDF MALWARE PADA LAYANAN AGREGATOR GARBA RUJUKAN DIGITAL (GARUDA) DENGAN ALGORITMA DECISION TREE

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

OLEH :

NOVI YUNINGSIH
09011281823133

JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA

2022

HALAMAN PENGESAHAN

DETEKSI ANOMALI FILE PDF MALWARE PADA LAYANAN AGREGATOR GARBA RUJUKAN DIGITAL (GARUDA) DENGAN ALGORITMA DECISION TREE

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh

Novi Yuningsih
09011281823133

Indralaya, Desember 2022

Pembimbing I Tugas Akhir	Pembimbing II Tugas Akhir
Deris Stiawan, M.T., Ph.D.	**Tri Wanda Septian, M.Sc.**
NIP. 197806172006041002	**NIK. 1901062809890001**

Mengetahui, 29/12/22
Ketua Jurusan Sistem Komputer

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

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

Hari : Jum'at

Tanggal : 18 November 2022

Tim Penguji :

1. **Ketua : Ahmad Fali Oklilas, S.T., M.T.**
2. **Sekretaris : Abdurrahman, S.Kom., M.Han.**
3. **Penguji : Huda Ubaya, M.T.**
4. **Pembimbing 1 : Deris Stiawan, M.T., Ph.D.**
5. **Pembimbing 2 : Tri Wanda Septian, M.Sc.**

Mengetahui,

Ketua Jurusan Sistem Komputer

Dr. Ir. H. Sukemi, M.T.

NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Novi Yuningsih

NIM : 09011281823133

Judul : Deteksi Anomali File PDF Malware pada Layanan Agregator Garba Rujukan Digital (GARUDA) dengan Algoritma Decision Tree

Hasil Pengecekan Software iThenticate/Turnitin : 3%

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.

Indralaya, Desember 2022

Novi Yuningsih

NIM. 09011281823133

HALAMAN PERSEMBAHAN

“Tugas Akhir ini kupersembahkan untuk Kedua Orang Tua-ku,
yang telah banyak memberikan dukungan dan semangat serta do’a yang
tidak pernah putus sehingga aku dapat menyelesaikan studi-ku ini,
Kalian sangatlah berharga bagiku, Bapak dan Mamak.

Dan ter-untuk diriku sendiri,
terima kasih karena telah melakukan yang terbaik.”

فَإِنَّ مَعَ ٱلْعُسْرِ يُسْرًا
إِنَّ مَعَ ٱلْعُسْرِ يُسْرًا

“*Maka sesungguhnya bersama kesulitan ada kemudahan,*
sesungguhnya bersama kesulitan ada kemudahan.”
(QS. Al-Insyirah: 5-6)

“Kalau terlalu sulit menjadi orang yang pintar, jadilah saja orang yang baik.”
(Reigen Arataka – Mob Psycho)

“때때로 불행한 일이 좋은 사람들에게 생길수 있다.”
“Bad things at times do happen to good people.”
(Hospital Playlist 2)

KATA PENGANTAR

بِسْـــــــــــمِ اللهِ الرَّحْمَنِ الرَّحِيْمِ

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji syukur Alhamdulillah penulis panjatkan atas kehadirat Allah SWT yang telah memberikan karunia dan rahmat-Nya, sehingga penulis dapat menyelesaikan penulisan Tugas Akhir ini dengan judul **"Deteksi Anomali File PDF Malware pada Layanan Agregator Garba Rujukan Digital (GARUDA) dengan Algoritma Decision Tree"**.

Dalam Tugas Akhir ini penulis menjelaskan bagaimana pemodelan untuk melakukan deteksi dan klasifikasi file *PDF Malware* pada dataset GARUDA. Penulis berharap kedepannya tulisan ini dapat digunakan bagi orang banyak dan menjadi bahan bacaan yang bermanfaat pada bidang *Network Security.*

Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada beberapa pihak atas ide dan saran serta bantuannya dalam menyelesaikan penulisan Tugas Akhir ini. Oleh karena itu, penulis menyampaikan ucapan rasa hormat dan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT, yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penulisan Tugas Akhir ini dengan baik dan lancar.
2. Orang Tua Tercinta, yang selalu mendoakan dan memberikan semangat serta dukungannya baik moril maupun materil selama ini.
3. Bapak Jaidan Jauhari, M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Dr. Ir. H. Sukemi, M.T., selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Bapak Sutarno, S.T., M.T. selaku Pembimbing Akademik penulis di Jurusan Sistem Komputer.
6. Bapak Deris Stiawan, M.T., Ph.D., selaku Dosen Pembimbing I Tugas Akhir yang telah memberikan bimbingan dalam menyelesaikan Tugas Akhir ini.
7. Kak Tri Wanda Septian, M.Sc., selaku Dosen Pembimbing II Tugas Akhir dan Mbak Nurul Afifah, M.Kom., yang telah bersedia meluangkan waktunya guna memberikan saran, motivasi, dan membimbing penulis dengan sangat baik.

8. Rekan-rekan riset Malware Geeks Infosec Unsri yaitu Rani, Alifah, Indah, Shena, dan Nata yang telah berjuang bersama sehingga dapat menyelesaikan riset yang begitu luar biasa ini.
9. Kakak, Mbak-mbak dan adik yang telah banyak membantu dan memberikan semangat kepada penulis dalam menyelesaikan perkuliahan ini.
10. Keponakan-keponakan penulis yang sangat lucu yaitu Pina, Ezi, dan Faruq yang selalu menjadi *mood booster* terbaik.
11. Teman-teman terdekat penulis, Ades, Nia, dan Rahma yang telah berbagi suka duka dan canda tawa serta terima kasih atas segala pengertian dan kebersamaan kalian selama ini.
12. Teman-teman seperjuangan Jurusan Sistem Komputer Angkatan 2018.
13. Dan seluruh pihak yang telah membantu serta memberikan semangat dan do'a.

Penulis menyadari bahwa penulisan Tugas Akhir ini masih sangat jauh dari kata sempurna. Untuk itu kritik dan saran yang membangun sangatlah diharapkan penulis agar dapat segera diperbaiki sehingga laporan ini bisa dijadikan sebagai masukkan ide dan pemikiran yang bermanfaat bagi semua pihak dan menjadi tambahan bahan bacaan bagi yang tertarik dalam penelitian pada bidang *Network Security.*

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Indralaya, Desember 2022
Penulis,

Novi Yuningsih
NIM. 09011281823133

DETEKSI ANOMALI FILE PDF MALWARE PADA LAYANAN AGREGATOR GARBA RUJUKAN DIGITAL (GARUDA) DENGAN ALGORITMA DECISION TREE

NOVI YUNINGSIH (09011281823133)
Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya
Email: noviiyuningsih@gmail.com

ABSTRAK

Format Dokumen Portabel (PDF) merupakan sebuah media pertukaran dokumen yang sangat rentan oleh serangan berbahaya yakni serangan PDF Malware. Salah satu layanan yang paling sering menggunakan file PDF sebagai medianya ialah layanan publikasi ilmiah yaitu Garba Rujukan Digital (GARUDA). Maka dari itu dilakukan penelitian menggunakan metode analisis statis terhadap masing-masing file PDF dan dilakukan ekstraksi data menggunakan PDFiD. Berdasarkan penelitian tersebut ditemukan sebuah keanehan atau anomali terhadap beberapa file PDF sehingga dataset dibagi menjadi tiga kelas yaitu *PDF benign, PDF anomaly*, dan *PDF malware*. Dataset yang dihasilkan pada penelitian ini adalah dataset dengan kondisi tidak seimbang dan digunakan *Synthetic Minority Oversampling Technique* (SMOTE) dan *NearMiss* untuk menyeimbangkan data. Untuk mengklasifikasikan serangan file PDF Malware digunakan salah satu metode *machine learning* yang cukup terkenal yaitu Algoritma Decision Tree. Proses klasifikasi dilakukan menjadi dua jenis yaitu klasifikasi dengan dataset asli (tidak seimbang) dan klasifikasi dengan dataset seimbang. Kemudian untuk memvalidasi keakuratan model klasifikasi, digunakan metode *cross validation* yaitu *Stratified K-Fold Cross Validation*. Berdasarkan hasil klasifikasi, performa yang paling baik didapatkan dengan persentase rata-rata nilai akurasi sebesar 99,83%, presisi 99,83%, recall 99,83%, *F1-score* 99,84%, TNR (*true negative rate*) 99,92%, AUC (*area under curve*) 99,88%, dan FPR (*false positive rate*) 0,001 serta FNR (*false negative rate*) 0,002.

Kata Kunci : PDF Malware, Deteksi Anomali, Multi-kelas, *Synthetic Minority Oversampling Technique* (SMOTE), *Stratified K-Fold Cross Validation*, Algoritma Decision Tree

MALWARE PDF FILES ANOMALY DETECTION ON GARBA RUJUKAN DIGITAL (GARUDA) AGREGATOR SERVICE USING DECISION TREE ALGORITHM

NOVI YUNINGSIH (09011281823133)
Computer Engineering Department, Computer Science Faculty, Sriwijaya University
Email : noviiyuningsih@gmail.com

ABSTRACT

Portable Document Format (PDF) is a document exchange media that is very vulnerable to malicious attacks, namely Malware PDF. One of the services that most often use PDF files as a medium is a scientific publication service Garba Rujukan Digital (GARUDA). Therefore, research was conducted using static analysis methods for each PDF and data extraction using PDFiD. Based on these research, it found an oddity or anomaly to some PDF files so that the dataset is divided into three classes, namely PDF benign, PDF anomaly, and PDF malware. The generated dataset in this research is a dataset with imbalanced conditions and used Synthetic Minority Oversampling Technique (SMOTE) and NearMiss to balance the data. To classify malware PDF file attacks used one of the well-known machine learning methods, Decision Tree Algorithm. Classification divided into two types, classification with the original dataset (imbalanced dataset conditions) and classification with balanced dataset conditions. Then to validate the accuracy of the classification model used cross validation method, Stratified K-Fold Cross Validation. Based on classification results, the best performance obtained by the average percentage of accuracy 99.83%, precision 99.83%, recall 99.83%, F1-score 99.84%, TNR (true negative rate) 99.92%, AUC (area under curve) 99.88%, and FPR (false positive rate) 0.001 and FNR (false negative rate) 0.002.

Keywords *: Malware PDF, Anomaly Detection, Multi-class, Synthetic Minority Oversampling Technique (SMOTE), Stratified K-Fold Cross Validation, Decision Tree Algorithm*

DAFTAR ISI

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR TABEL

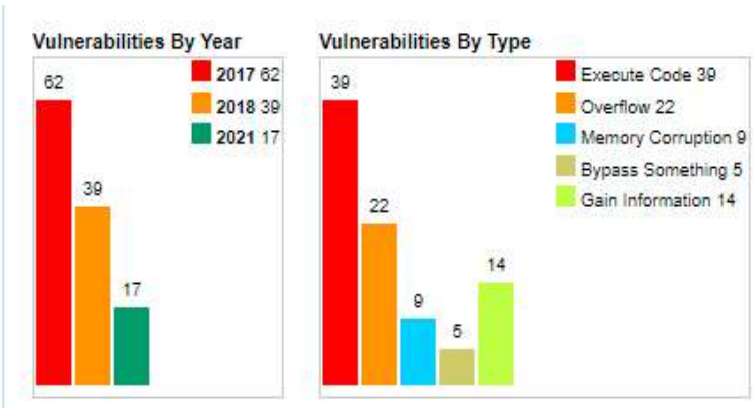
Halaman

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Portable Document Format atau yang lebih umum dikenal sebagai PDF adalah salah satu bentuk format dokumen yang banyak dijumpai pada beberapa dekade terakhir. Format Dokumen Portabel digunakan sebagai standar *defacto* untuk berbagi dokumen [1]. Kondisi ini disebabkan oleh karakteristik PDF seperti fleksibilitas dan portabilitas dalam lintas platform. Penggunaan *Portable Document Format* sebagai format deskripsi dokumen yang paling luas dan digunakan di seluruh dunia juga dipengaruhi oleh bahasa pemrograman yang dimiliki oleh PDF itu sendiri, sangat didedikasikan untuk pembuatan dan manipulasi dokumen yang telah mengumpulkan banyak fitur pemograman yang canggih dari versi ke versi [2]. Terdapat miliaran file PDF tersebar diinternet dan tidak semua file PDF tersebut merupakan file PDF yang aman untuk digunakan. Terkadang tanpa disadari, terdapat beberapa file PDF diinternet yang berbahaya dimana secara sengaja atau tidak sengaja dapat mencoba mengeksploitasi untuk menginfeksi sebuah mesin atau komputer seseorang. Sebuah file PDF memungkinkan dapat berisi berbagai objek, seperti kode JavaScript atau kode biner [3].



Gambar 1.1 Hasil Statistik CVE Details [4]

Berdasarkan CVE Details, hingga tahun 2021 telah ditemukan kurang lebih sebanyak 118 kerentanan pada Adobe Acrobat Reader [4]. Setiap pembaca file PDF memiliki kerentanannya masing-masing dan sebuah file PDF yang

berbahaya dapat menemukan cara untuk memanfaatkan celah dari kerentanan tersebut.

Garba Rujukan Digital atau GARUDA merupakan sebuah platform atau layanan sumber informasi publikasi ilmiah yang ada di Indonesia dimana layanan ini dikelola oleh Kemenristekdikti. Dalam layanan GARUDA dapat mencakup semua aspek ilmu pengetahuan seperti seni, humaniora, ilmu perilaku, ilmu sosial, fisika, teknik, matematika dan komputer, kimia dan biologi, dan lain sebagainya [5]. Awal mula nama GARUDA ialah RII (Referensi Ilmiah Indonesia) [6]. GARUDA adalah salah satu layanan yang berisikan referensi publikasi ilmiah di Indonesia dan memberikan akses terhadap karya ilmiah yang dihasilkan oleh akademisi dan peneliti Indonesia. Hingga saat ini, kurang lebih terdapat 2.700 *publisher* dan 14.000 *journal* yang dapat diakses pada layanan ini [7].

Sebuah *dataset* atau sumber data pada umumnya memiliki nilai-nilai pada setiap objek yang tidak terlalu berbeda jauh antara objek yang satu dengan objek yang lain. Namun terkadang, pada *dataset* tersebut juga dapat ditemukan objek-objek yang mempunyai nilai atau sifat atau karakteristik yang berbeda dibandingkan dengan objek pada umumnya [8]. Dalam sebuah penelitian, objek yang tidak normal dalam sebuah dataset dapat menurunkan hasil performansi. Maka dari itu, pendeteksian anomali merupakan sebuah proses yang dapat dilakukan untuk mengidentifikasi objek yang tidak normal dalam sebuah kumpulan data. Proses menemukan pola objek dalam sebuah dataset yang perilakunya tidak normal disebut dengan Deteksi Anomali [9]. Sebutan lain untuk perilaku objek yang tidak normal tersebut ialah anomali atau *outlier*.

Berdasarkan penelitian yang telah dilakukan oleh Nandhini dan Dr. Jeen Marseline pada tahun 2020 mengenai evaluasi performa algoritma-algoritma *machine learning* pada deteksi spam email yang mana memiliki kesamaan jenis data yang digunakan pada penelitian ini yaitu data numerik dan memperoleh hasil akurasi tertinggi yaitu sebesar 99,93%. Pada penelitian tersebut, algoritma yang memperoleh akurasi tertinggi menggunakan Algoritma Decision Tree [10]. Metode Decision Tree merupakan sebuah metode klasifikasi yang dapat mengubah kenyataan (fakta) yang sangat besar menjadi sebuah pohon keputusan (*decision tree*) yang menggambarkan aturan-aturan atau *rules*. Pohon keputusan ini juga

dapat digunakan untuk mengeksplorasi data, serta dapat menemukan hubungan tersembunyi antara sejumlah calon variabel *input* dengan sebuah variabel target. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, algoritma tersebut antara lain yaitu ID3, C4.5, CART [11]. Beberapa subjek yang telah disebutkan diatas merupakan hal-hal yang melatar-belakangi penulis dalam mengambil judul "Deteksi Anomali File PDF Malware pada Layanan Agregator Garba Rujukan Digital (GARUDA) dengan Algoritma Decision Tree" dikarenakan metode Decision Tree dianggap memiliki tingkat akurasi yang baik dalam mengklasifikasikan sebuah data.

1.2 Rumusan Masalah

Adapun rumusan masalah dalam penulisan Tugas Akhir ini ialah sebagai berikut.

1. Bagaimana teknik yang digunakan untuk mengekstrak *dataset* file PDF Malware GARUDA (Garba Rujukan Digital)?
2. Bagaimana cara melakukan klasifikasi serangan file PDF Malware pada dataset GARUDA (Garba Rujukan Digital)?
3. Bagaimana cara mengevaluasi performa Algoritma Decision Tree dalam mendeteksi serangan file PDF Malware*?*

1.3 Batasan Masalah

Batasan atau ruang lingkup masalah pada Tugas Akhir ini ialah sebagai berikut.

1. Penelitian yang dilakukan mencakup permasalahan serangan file PDF Malware.
2. Algoritma yang digunakan untuk mendeteksi serangan adalah Algoritma Decision Tree.
3. *Dataset* yang digunakan berasal dari server Layanan Agregator Garba Rujukan Digital (GARUDA).
4. Tidak membahas cara pencegahan serangan file PDF Malware.

1.4 Tujuan

Adapun tujuan yang ingin dicapai dalam penulisan dan penelitian pada Tugas Akhir ini ialah sebagai berikut.

1. Mengekstraksi *dataset* file PDF Malware GARUDA (Garba Rujukan Digital) menggunakan teknik analisis malware yaitu analisis statis.
2. Mengklasifikasikan serangan file PDF Malware pada *dataset* GARUDA (Garba Rujukan Digital) menggunakan Algoritma Decision Tree.
3. Mengevaluasi performa Algoritma Decision Tree dalam mendeteksi serangan file PDF Malware menggunakan *Stratified K-Fold Cross Validation* dan *Confusion Matrix*.

1.5 Manfaat

Adapun manfaat yang dapat diambil dari penulisan Tugas Akhir ini ialah sebagai berikut.

1. Dapat menerapkan teknik analisis malware secara statis untuk mengubah data file PDF menjadi angka.
2. Dapat menerapkan algoritma *Synthetic Minority Oversampling Technique* (SMOTE) dan *NearMiss* untuk menyeimbangkan kondisi dataset.
3. Dapat menerapkan metode Decision Tree untuk mengklasifikasikan serangan file PDF Malware dalam kasus data *multiclass*.
4. Dapat membantu mempermudah penelitian mengenai serangan file PDF Malware pada layanan publikasi ilmiah.

1.6 Metodelogi Penelitian

Adapun metodologi penelitian yang digunakan dalam penulisan Tugas Akhir ini akan melewati beberapa tahapan seperti berikut.

1. Metode Studi Pustaka dan Studi Literatur

   Metode ini dilakukan dengan cara mencari dan mengumpulkan referensi yang berupa studi literatur terhadap *paper* atau jurnal yang terdapat pada buku dan internet mengenai “Deteksi Anomali File PDF Malware dengan Algoritma Decision Tree”.

2. Metode Konsultasi

   Metode konsultasi ini dilakukan dengan cara konsultasi kepada pihak-pihak yang memiliki pengetahuan serta wawasan yang baik dalam mengatasi permasalahan yang ditemui pada penulisan tugas akhir "Deteksi Anomali File PDF Malware dengan Algoritma Decision Tree".

3. Metode Perancangan Model

   Dalam metode perancangan model ini dilakukan dengan cara membuat sebuah perancangan atau pemodelan yang mana nantinya akan disimulasikan dalam bentuk program atau simulasi program.

4. Metode Pengujian

   Metode pengujian ini dilakukan dengan cara melakukan uji coba terhadap simulasi program yang telah dibuat pada metode sebelumnya. Metode ini dilakukan dengan tujuan mengetahui hasil dari simulasi program tersebut dapat menghasilkan performa yang baik atau tidak. Performa dapat ditunjukkan menggunakan nilai akurasi, nilai presisi, dan nilai *error rate*.

5. Metode Analisa dan Kesimpulan

   Pada metode analisa dan kesimpulan ini akan dilakukan analisis terhadap hasil dari pengujian yang mana mencakup kelebihan dan kekurangan penelitian sehingga hasilnya dapat digunakan untuk penelitian-penelitian yang berkaitan.

1.7 Sistematika Penulisan

Sub-bab sistematika penulisan dapat digunakan untuk menjelaskan dan menegaskan bab-bab yang akan dituliskan pada Tugas Akhir ini. Adapun sistematika penulisan yang dibuat dalam Tugas Akhir ini ialah sebagai berikut.

BAB I PENDAHULUAN

Pada Bab I akan menjabarkan mengenai latar belakang permasalahan yang diambil dalam penelitian, perumusan masalah, pembatasan masalah, tujuan dan manfaat penelitian, metodologi penelitian, dan sistematika penulisan Tugas Akhir.

BAB II TINJAUAN PUSTAKA

Pada Bab II akan dijabarkan mengenai sumber topik dan literatur serta teori-teori mana saja yang diambil yang berkaitan dengan serangan PDF Malware, SMOTE *(Synthetic Minority Oversampling Technique), Stratified K-Fold Validation*, dan Algoritma Decision Tree.

BAB III METODOLOGI PENELITIAN

Pada Bab III ini akan dijelaskan mengenai langkah-langkah atau metode yang digunakan yaitu deteksi anomali serangan PDF Malware dengan Algoritma Decision Tree secara *step-by-step* yang diaplikasikan dengan menggunakan blok diagram, *flowchart,* dan yang lainnya.

BAB IV HASIL DAN PEMBAHASAN

Pada Bab IV ini akan dijabarkan mengenai hasil dari penelitian yang telah dilakukan yaitu deteksi anomali serangan PDF Malware dengan Algoritma Decision Tree yang diterapkan secara rinci dan detail.

BAB V KESIMPULAN DAN SARAN

Bab V ini akan diisi menggunakan kesimpulan yang dapat diambil dari setiap bab yang telah disusun sebelumnya mengenai hasil dari implementasi Algoritma Decision Tree dalam deteksi anomali serangan PDF Malware. Pada bab ini juga akan diisi dengan saran yang mana penulis berharap akan dapat berguna untuk penelitian selanjutnya yang berkaitan dengan penelitian Tugas Akhir ini.

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR PUSTAKA

[1] H. V. Nath and B. M. Mehtre, "Ensemble Learning For Detection Of Malicious Content Embedded In PDF Documents," *2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems*, pp. 1-4, 2015.

[2] E. Filiol, A. Blonce, and L. Frayssignes, "Portable Document Format (PDF) Security Analysis And Malware Threats," *Presentations of Europe BlackHat 2008 Conference, Amsterdam*, pp. 2–8, 2008.

[3] B. Cuan, A. Damien, C. Delaplace, and M. Valois, "Malware Detection in PDF Files Using Machine Learning," *ICETE 2018 - Proceedings of the 15th International Joint Conference on e-Business and Telecommunication*, vol. 2, pp. 412–419, 2018.

[4] D. CVE, "Vulnerability Statistics-Adobe Acrobat Reader," *CVE Details*, 2021. https://www.cvedetails.com/product/32069/Adobe-Acrobat-Reader-Dc.html?vendor_id=53 (accessed Apr. 20, 2022).

[5] Pustaka IAINBukitTinggi, "Tutorial Akses GARUDA (Garba Rujukan Digital)," 2018. https://pustaka.iainbukittinggi.ac.id/pustaka/2680/tutorial-akses-garuda-garba-rujukan-digital/ (accessed Apr. 15, 2022).

[6] R. Wahyudin, "Garuda - Garba Rujukan Digital," *Jurnal Pustakan Indonesia*, vol. 10, no. 1, pp. 62–63, 2010.

[7] GARUDA, "Home-Garba Rujukan Digital," *Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi*, 2022. https://garuda.kemdikbud.go.id/ (accessed Apr. 20, 2022).

[8] M. A. Bijaksana and E. B. Setiawan, "Analisis Perbandingan Clustering-Based, Distance-Based dan Density-Based dalam Mendeteksi Outlier Social Network Analysis View project," *Seminar Nasional Aplikasi Teknologi Informasi (SNATI 2009)*, pp. 101–108, 2009.

[9] S. Agrawal and J. Agrawal, "Survey On Anomaly Detection Using Data

Mining Techniques," *Procedia Computer Science*, vol. 60, no. 1, pp. 708–713, 2015.

[10] S. Nandhini and K. S. Dr. Jeen Marseline, "Performance Evaluation of Machine Learning Algorithms for Bitcoin Price Prediction," *Proceedings of the 4th International Conference on Inventive Systems and Control, ICISC 2020*, pp. 110–114, 2020.

[11] F. Dwi Meliani Achmad, Budanis, Slamat, "Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree," *Jurnal IPTEK*, vol. 16, no. 1, pp. 18–23, 2012.

[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," *Jurnal Online Informatika*, vol. 3, no. 2, p. 99, 2019.

[13] N. Fleury, T. Dubrunquez, and I. Alouani, "PDF-Malware: An Overview on Threats, Detection and Evasion Attacks," 2021.

[14] N. Srndic and P. Laskov, "Detection of Malicious PDF Files Based on Hierarchical Document Structure," *Proceedings of the 20th Annual Network & Distributed Systems Symposium*, 2013.

[15] F. Edition, "Document management — Portable document format — Part 1: PDF 1.7," *Adobe Systems Incorporated 2008 – All rights reserved*, pp. 1-748, 2008.

[16] J. S. Cross and M. A. Munson, "Deep PDF Parsing to Extract Features for Detecting Embedded Malware," no. 9, pp. 1–18, 2011.

[17] Y. Sun *et al.*, "MMPD : A Novel Malicious PDF File Detector for Mobile Robots," no. 3, pp. 1–11, 2020.

[18] D. Stevens, "PDFiD," 2009. https://blog.didierstevens.com/2009/03/31/pdfid/ (accessed Mar. 06, 2022).

[19] R. Chalapathy and S. Chawla, "Deep Learning For Anomaly Detection: A Survey," pp. 1–50, 2019.

[20] C. Callegari *et al.*, “A Methodological Overview On Anomaly Detection,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7754, pp. 148–183, 2013.

[21] M. Ahmed, A. Naser Mahmood, and J. Hu, “A Survey Of Network Anomaly Detection Techniques,” *Journal of Network and Computer Applications*, vol. 60, pp. 19–31, 2016.

[22] C. C. Aggarwal, "Outlier Analysis Second Edition," *Springer Nature,* vol. 2, pp. 1-463, 2017.

[23] V. Chandola, U. Vipin, and A. Banerjee, “Anomaly Detection: A Survey,” *Computers, Materials and Continua*, vol. 14, no. 1, pp. 1–22, 2009.

[24] H. He, W. Zhang, and S. Zhang, “A Novel Ensemble Method For Credit Scoring: Adaption Of Different Imbalance Ratios,” *Expert Systems with Applications*, vol. 98, pp. 105–117, 2018.

[25] A. R. B. Alamsyah, S. Rahma, N. S. Belinda, and A. Setiawan, “SMOTE and Nearmiss Methods for Disease Classification with Unbalanced Data,” pp. 305–314, 2021.

[26] N. M. Mqadi, N. Naicker, and T. Adeliyi, “Solving Misclassification of the Credit Card Imbalance Problem Using Near Miss,” *Mathematical Problems in Engineering.*, vol. 2021, 2021.

[27] J. Li, S. Fong, and Y. Zhuang, “Optimizing SMOTE by Metaheuristics with Neural Network and Decision Tree,” 2015.

[28] W. Xie, G. Liang, Z. Dong, B. Tan, and B. Zhang, “An Improved Oversampling Algorithm Based on the Samples Selection Strategy for Classifying Imbalanced Data,” *Mathematical Problems in Engineerin*g., vol. 2019, 2019.

[29] T. E. Tallo and A. Musdholifah, “The Implementation of Genetic Algorithm in Smote (Synthetic Minority Oversampling Technique) for Handling Imbalanced Dataset Problem,” *Proceedings - 2018 4th International*

Conference on Science and Technology, ICST 2018, vol. 1, pp. 1–4, 2018.

[30] Y. T. Samuel, C. Beatrix, and A. Nahuway, "Prediksi Indeks Prestasi Mahasiswa Yang Berkuliah Sambil Bekerja Di Universitas Advent Indonesia Dengan Menggunakan Metode Decision Tree C4.5 Dan SMOTE," pp. 69–77, 2020.

[31] I. Sutoyo, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik," vol. 14, no. 2, 2018.

[32] V. S. Ginting, K. Kusrini, and E. Taufiq, "Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python," *Jurnal Teknologi Informatika dan Komunikasi*, vol. 10, no. 1, pp. 36–44, 2020.

[33] Nurhayati, I. Soekarno, I. K. Hadihardaja, and M. Cahyono, "A Study Of Hold-Out And K-Fold Cross Validation For Accuracy Of Groundwater Modeling In Tidal Lowland Reclamation Using Extreme Learning Machine," *Proceedings of 2014 2nd International Conference on Technology, Informatics, Management, Engineering and Environment, TIME-E 2014*, pp. 228–233, 2015.

[34] C. Ulucenk, "Techniques for Analysing PDF Malware," no. 12, 2011.