VISUALISASI DAN KLASIFIKASI *MALWARE* MENGGUNAKAN METODE *RANDOM FOREST*

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Komputer

Oleh

Andre Ghazali Armi
09011281520099

JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2021

LEMBAR PENGESAHAN

VISUALISASI DAN KLASIFIKASI *MALWARE* MENGGUNAKAN METODE *RANDOM FOREST*

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

ANDRE GHAZALI ARMI
09011281520099

Indralaya, Januari 2021

Mengetahui,
Ketua Jurusan Sistem Komputer

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

Pembimbing

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

HALAMAN PERSETUJUAN

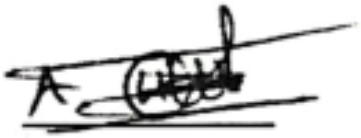
Telah diuji dan lulus pada :

Hari : Kamis
Tanggal : 31 Desember 2020

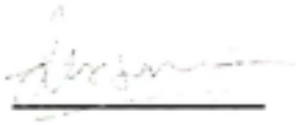
Tim Penguji :

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

2. Anggota I : Ahmad Heryanto, M.T.



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



Mengetahui,
Ketua Jurusan Sistem Komputer

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

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Andre Ghazali Armi

NIM : 09011281520099

Judul : VISUALISASI DAN KLASIFIKASI MALWARE MENGGUNAKAN METODE RANDOM FOREST

Hasil Pengecekan *Software iThenticate/Turnitin* : 7%

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

Demikian, pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan.

Palembang, Januari 2021

Andre Ghazali Armi

NIM. 09011281520099

HALAMAN PERSEMBAHAN

Skripsi ini Saya Persembahkan untuk :

Maha Agung,Allah Subhanahu wa Ta'ala
Untuk Mamakku Mutia
Wanita Nomor 1 di dunia
Untuk Papaku Dt. Rajo Imbang
Lelaki Panutanku dan Penuh Tanggung Jawab
Esok Lusa Kita Akan Bertemu Kembali
Opung, Uwak, Ocik dan Tulang
Kakak, Abang, Maodang dan Paodang
Seluruh Teman Kosan
Keluarga Sistem Komputer
Fakultas Ilmu Komputer
Universitas Sriwijaya
Indonesia
Bumi
Alam Semesta

"goodluck y'all and kthxbye"

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji dan syukur penulis panjatkan kehadirat Allah SWT, atas segala karunia dan rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan Proposal Tugas Akhir ini dengan judul "**Visualisasi dan Klasifikasi Malware Menggunakan Metode Random Forest**".

Dalam laporan ini penulis menjelaskan mengenai Visualisasi malware mengklasifikasikannya dengan *Random Forest* . Penulis berharap tulisan ini dapat bermanfaat bagi orang banyak, dan menjadi tambahan bahan bacaan bagi yang tertarik meneliti tentang pengklasifikasian malware.

Pada penyusunan proposal tugas akhir ini, tidak terlepas dari bantuan, bimbingan serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis mengucapkan rasa syukur dan terima kasih kepada yang terhormat :

1. Allah SWT, yang telah memberikan kemudahan, kesehatan, serta kesempatan dalam pelaksanaan pembuatan Tugas Akhir ini.
2. Mamak, Papa dan semua keluarga besar yang telah memberikan dukungan dan nasehat-nasehat serta motivasi selama ini. Terima kasih atas dukungan baik berupa moral, material, maupun spiritual.
3. Bapak Jaidan Jauhari, S.Pd., M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya
4. Bapak Dr. Ir. H. Sukemi, M.T. selaku Ketua Jurusan Sistem Kompuer Fakutas Ilmu Komputer Universitas Sriwijaya
5. Bapak Deris Stiawan, M.T., PH.D. selaku Pembimbing Tugas Akhir Penulis di Jurusan Sistem Komputer. Terima kasih karena telah meluangkan waktunya untuk membimbing penulis dalam menyelesaikan tugas akhir ini serta telah memberikan bimbingan dan nasehat selama perkuliahan.
6. Bapak Deris Stiawan, M.T., PH.D. selaku Dosen Pembimbing Akademik

7. Seluruh Dosen Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya.
8. Teman-teman seperjuangan Sistem Komputer angkatan 2015 dan anak-anak SKC khususnya yang selalu bersama selama perkuliahan ini .
9. Serta semua pihak yang telah membantu baik moril maupun materil yang tidak dapat disebutkan satu persatu dalam penyelesaian tugas akhir ini. Terima kasih banyak semuanya.

Penulis menyadari bahwa masih terdapat banyak kekurangan dalam penulisan Tugas Akhir ini, baik dari materi maupun teknik penyajiannya, mengingat kurangnya pengetahuan dan pengalaman penulis. Untuk itu, penulis mengharapkan adanya kritik dan saran yang membangun agar dapat memperbaiki kekurangan-kekungan tersebut kedepannya nanti.

Akhir kata dengan segala keterbatasan, penulis berharap semoga penulisan Tugas Akhir ini dapat menjadi tambahan wawasan dan ilmu pengetahuan bagi mahasiswa yang memerlukan khususnya mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya secara langsung ataupun tidak langsung sebagai sumbangan pikiran dalam peningkatan mutu pembelajaran.

Palembang, Januari 2021

Penulis

MALWARE VISUALIZATION AND CLASIFICATION USING RANDOM FOREST METHOD

Andre Ghazali Armi (09011281520099)

Dept of Computer Engineering, Faculty of Computer Science,

Sriwijaya University

Email: andreghazaliarmi@gmail.com

Abstract

Visualization has a function so that we can see the malware in grayscale form which consists of data in the form of a collection of hexadecimal numbers which are converted into decimal. Malware classification is a way to identify and classify malware based on their respective groups. Random forest is one of the many classification methods used for this case. Local Binary pattern used for feature extraction process from existing data. The system is trained and tested using 1000 data from 10 different family malware with a comparison of 8: 2 training and test data. In this study, we utilized an approach of converting a malware binary into an image and use Random Forest to classify various malware families. The resulting accuracy of 0.99-0.995 exhibits the effectivess of the method in detecting malware

Keywords : *Malware, Grayscale, Local Binary Pattern, Random Forest, Clasification*

VISUALISASI DAN KLASIFIKASI *MALWARE* MENGGUNAKAN METODE *RANDOM FOREST*

Andre Ghazali Armi (09011281520099)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email: *andreghazaliarmi@gmail.com*

Abstrak

Visualisasi di lakukan agar kita dapat melihat malware dalam bentuk grayscale yang terdiri dari data berupa kumpulan angka hexadecimal yang dirubah menjadi decimal. Klasifikasi malware merupakan salah satu cara untuk mengidentifikasi dan mengelompokan malware berdasarkan kelompoknya masing masing. Random forest adalah salah satu dari sekian banyak metode klasifikasi yang digunakan kali ini. Local Binary pattern digunakan untuk proses ekstraksi fitur dari data yang ada. Sistem dilatih dan diuji menggunakan 1000 data dari 10 family malware yang berbeda dengan perbandingan data latih dan uji 8 : 2. Dalam studi ini, kami menggunakan pendekatan untuk mengubah biner malware menjadi gambar dan menggunakan Random Forest untuk mengklasifikasikan berbagai keluarga malware. Akurasi yang dihasilkan sebesar 0,99-0.995 menunjukkan keefektifan metode dalam mendeteksi malware

Kata Kunci : *Malware, Grayscale, Local Binary Pattern, Random Forest,* Klasifikasi

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I
PENDAHULUAN

1.1 Latar Belakang

Pada zaman sekarang ini, perkembangan Malware sangat pesat dan menjadi sorotan bagi penggiat teknologi di berbagai penjuru dunia, malware sampai sekarang ini selalu menjadi ancaman bagi para penggiat internet dan server dikarenakan selalu menyerang keamanan internet [1]. Awalnya malware dibentuk untuk menyerang user dan melakukan Tindakan illegal pada yang menjurus ke cyber crime seperti merusak data , mencuri data , membebani sistem , bahkan sekarang ini di gunakan sebagai protes kepada pemerintah ataupun saingan bisnis. Beberapa contoh tersebut terjadi di dunia nyata dan bukan sekedar kabar angin yang bisa dianggap lelucon dan sepele. Dikarenakan resiko itu maka sangat banyak penggiat IT untuk mencegah dan memperkecil dampak yang ditimbulkan dari Malware.

Hal ini menjadikan para penggiat anti malware agar selalu update akan proteksi untuk menahan serangan dari berbagai virus dan malware baru yang ada sekarang ini, adar dapat menyajikan rasa aman dan nyaman kepada pengguna. Perkembangan malware pada zaman ini sangat berpengaruh pada sekarang karena data yang semakin banyak dan kecepatan pengolahan data yang cepat saat sekarang ini dibuat semakin rumit. Malware yang ada pada saat ini dapat menghindari deteksi dengan berbagai cara, seperti mengcloning diri nya di file lain, membuat file hiden, memproteksi diri nya dengan cara tidak bisa di hapus dan sebagainya [2].

Adapun klasifikasi malware dibutuhkan guna untuk mengelompokan malware pada family nya masing masing sesuai data yang di peroleh selama ini untuk menganalisa malware tersebut dengan berbagai cara yang ada , Adapun salah satu cara tersebut yaitu menjadikan malware kedalam bentuk grayscale lalu dikelompokkan menggunakan sebuah mesin yang dikenal sebagai machine learning agar sesuai nantinya dengan kelompok mereka nantinya [3].

Pada saat menganalisis malware yang ada mereka memberikan bentuk dan menampilkan dari malware itu dalam bentuk string biner yaitu 0 dan 1. File biner yaitu 8 bit yang dibentuk seperti 2D lalu di visualisasikan sebagai gambar hitam putih yaitu grayscale. Gambar ini nantinya digunakan untuk pendeteksi perubahan

perubahan yang kecil sembari dia mempertahankan bentuk dari segi besarnya. Agar nantinya disaat perubahan itu terlihat kita dapat dengan mudah secaara kasat mata melihat file nya pada bentuk grayscale

Algoritma Random Forest merupakan salah satu algoritma yang biasa digunakan untuk membantu pengklasifikasian data dalam jumlah yang besar,

1.2 Tujuan

Adapun tujuan dari penelitian ini adalah:

1. Memvisualisasikan binari malware ke dalam bentuk citra grayscale.
2. Memberikan penerapan dari algortima *Random Forest* untuk mengklasifikasikan family dari Malware.
3. Menganalisa akurat dan algoritma *Random Forest* dalam mengklasifikasi malware.

1.3 Manfaat

1. Dapat memahami algoritma *Random Forest* serta penerapan pada nantinya untuk mengklasifikasikan malware tersebut.
2. Agar digunakan sebagai referensi untuk penelitian nantinya .

1.4 Rumusan dan Batasan Masalah

1. Untuk penjelasan LBP pada fitur grayscale malware dan ekstraksi menggunakan LBP itu
2. Bagaimana akurasi sistem Random forest dalam pengklasifikasian malware yang nantinya awal di ubah pada citra grayscale

1.5 Metodologi Penelitian

Metode yang penulis gunakan pada tugas ini yaitu sebagai berikut untuk tahap tahapannya

1. Study Pustaka (Literatur)

   Tahap ini adalah tahap untuk awal dalam pencarian referensi dan pengangkatan judul berdasarkan judul yang nantinya akan dilakukan guna memberikan penunjang penelitian

2. Konsultasi

   Pada step ini penulis akan melakukan pertemuan dan perbincangan guna konsultasi kepada orang yang dianggap memiliki wawasan pada nanti tugas akhir

3. Pengumpulan Data

   Pada step ini tidak dilakukan pengumpulan data dikarenakan data yang akan digunakan pada penelitian ini adalah data yang sudah ada yaitu data dari CNNs dant BIG 2015

4. Pengolahan Data

   Pada step ini dilakukan mengolah data untuk memvisualisasikan malware serta menerapkan algoritma *Random Forest* dalam pengklasifikasian malware.

5. Analisa

   Padastep ini kita melakukan pengecekan data dan menvisualisasikan malware lalu menganalisis data tersebut.

6. Kesimpulan dan Saran

   Tahap ini adalah menarik kesimpulan dari Analisa dan hasil penelitian inidan di jadikan referensi nantinya untuk penelitian berikut nya .

1.6 Sistematika Penulisan

Dalam penulisan tugas akhir ini akan dibagi menjadi beberapa bagian bab dengan sistematika sebgai berikut :

BAB I. Pendahuluan

Pada bab pertama berisi latar belakang dan landasan terori kenapa penelitian ini di ambil dan alasaan yang ada pada tugas akhir ini ini.

BAB II. Tinjauan Pustaka

Pada bab kedua iniberisikan teori teroi dasar dari tugas akhir ini yang mana nanti berisikan LBP , *Malware , Feature extraction, Random Forrest.*

BAB III. Metodologi

Pada bab ketiga ini adalah dari penjelasan dari penelitian yang akan dilakukan serta step by step dari penelitian yang ada dari pengumpulan data proses dan cara dari penelitian ini .

BAB IV. Pengujian dan Analisa

Pada bab keempat ini adalah hasil dari penelitain dan pengujian yang dilakukan dan akan di analisis serta pembuktian yang sesuai dengan tujuan tugas akhir ini .

BAB V. Kesimpulan dan Saran

Pada bab kelima adalah kesimpulan beserta saran atas hasil penelitian kali ini

Daftar Pustaka

[1] Nataraj, L., Karthikeyan, S., Jacob, G., & Manjunath, B. S. (2011). Malware images: Visualization and automatic classification. ACM International Conference Proceeding Series.

[2] Ni, S., Qian, Q., & Zhang, R. (2018). Malware identification using visualization images and deep learning. Computers and Securit*y*, 77, 871–885.

[3] Songqing Yue. Imbalanced malware images classification: a CNN based approach. CoRR,abs/1708.08042, 2017. URL: http://arxiv.org/abs/1708.08042, arXiv:1708.08042

[4] Pascanu, R., Stokes, J. W., Sanossian, H., Marinescu, M., & Thomas, A. (2015). Malware classification with recurrent networks. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2015-August, 1916–1920.

[5] Makandar, A., & Patrot, A. (2017). Malware class recognition using image processing techniques. 2017 International Conference on Data Management, Analytics and Innovation, ICDMAI 2017, 76–80. https://doi.org/10.1109/ICDMAI.2017.8073489

[6] M. S. Alam and S. T. Vuong, Random Forest Classification for Detecting Android Malware, pp. 663–669, 2013.

[7]. Pullaperuma, P. P., & Dharmaratne, A. T. (2013). Taxonomy of file fragments using Gray-Level Co-Occurrence Matrices. 2013 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2013, 1–7. https://doi.org/10.1109/DICTA.2013.6691534

[8] L. E. O. Breiman, Random Forests, pp. 5–32, 2001.

[9] D. Ucci, L. Aniello, and R. Baldoni, Survey of machine learning techniques for malware analysis, Comput. Secur., vol. 81, pp. 123–147, 2019

[10] V. G. Biju, Kappa and Accuracy Evaluations of Machine Learning Classifiers, pp. 20–23, 2017.

[11] J. W. G. Putra, Pengenalan Konsep Pembelajaran dan Deep Learning, pp. 1–235, 2019

[12] A. H. Lashkari, A. F. A. Kadir, L. Taheri, and A. A. Ghorbani, Toward Developing a Systematic Approach to Generate Benchmark Android Malware Datasets and Classification, Proc. - Int. Carnahan Conf. Secur. Technol., vol. 2018-Octob, no. Cic, pp. 1–7, 2018.