VISUALISASI DAN KLASIFIKASI *MALWARE* MENGGUNAKAN METODE *K-NEAREST NEIGHBOR (K-NN)*

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
Jenjang S1

OLEH:

MEIDI DWI HAFIZ
09011281520097

JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA

2020

LEMBAR PENGESAHAN

VISUALISASI DAN KLASIFIKASI *MALWARE* MENGGUNAKAN METODE *K-NEAREST NEIGHBOR*

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

MEIDI DWI HAFIZ
09011281520097

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 :

Penguji : Ahmad Heryanto, M.T.

Mengetahui,
Ketua Jurusan Sistem Komputer,

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

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Meidi Dwi Hafiz

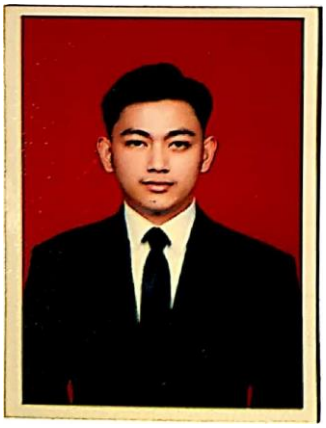
NIM : 09011281520097

Judul : VISUALISASI DAN KLASIFIKASI *MALWARE* MENGGUNAKAN METODE *K-NEAREST NEIGHBOR*

Hasil Pengecekan *Software iThenticate*/*Turnitin* : 20%

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

Meidi Dwi Hafiz
NIM. 09011281520097

HALAMAN PERSEMBAHAN

Tugas akhir ini saya persembahkan :

Yang Maha Kuasa, Allah Subhanahu wa Ta'ala
Mama dan Papa
Kedua adikku
Rekan seperjuangan skc 2015
Rekan 1 kos
Dan
Untuk yang selalu menanyakan kapan wisuda

Untuk saya

"Ada orang-orang berdoa siang malam untuk kamu. Jangan patahkan itu Jangam menyalahkan kondisi dan keadaan dengan buruknya manajemen kamu itu. Semoga kamu jadi lebih baik kedepan."

Terimakasih

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah SWT, karena berkat karunia dan rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan tugas akhir ini dengan judul "**Visualisasi dan Klasifikasi *Malware* Menggunakan Metode *K-Nearest Neighbor***".

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 terimakasih kepada yang terhormat :

1. Allah SWT yang karena atas kuasanya lah semua ilmu yang saya dapat dan tetap terus dalam keadaan sehat sehingga penyusunan tugas akhir ini dapat saya selesaikan.
2. Kedua orang tua saya dan adik beserta keluarga yang selalu mendoakan, memberi nasehat, serta memberi dukungan.
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 dan Pembimbing Akademik saya.
6. Seluruh Dosen Jurusan Sistem Komputer di Fakultas Ilmu Komputer Universitas Sriwijaya, terima kasih telah membimbing dan membagi ilmunya yang bermanfaat selama penulis mengikuti perkuliahan di Jurusan Sistem Komputer.
7. Teman-teman seperjuangan angkatan Sistem Komputer 2015 terkhusus Kelas SKC yang selalu membersamai selama perkuliahan dan teman-teman yang telah membantu saya dalam menyelesaikan tugas akhir ini.
8. Teman seperjuangan satu kos, yang selalu memotivasi dan perduli terhadap penulis dalam penyelesaian tugas akhir ini.

9. Almameter Universitas Sriwijaya yang telah memberi kesempatan dan fasilitas selama penulis menempuh pendidikan Strata 1 di Jurusan Sistem Komputer ini.

Penulis menyadari bahwa Laporan ini masih jauh dari kesempurnaan, oleh karena itu penulis sangat mengharapkan kritik dan saran yang bersifat membangun agar lebih baik lagi dikemudian hari.

Akhir kata dengan segala keterbatasan, penulis berharap semoga laporan ini menghasilkan sesuatu yang bermanfaat bagi kita semua khususnya bagi 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 CLASSIFICATION USING K-NEAREST NEIGHBOR METHOD

Meidi Dwi Hafiz (09011281520097)

Dept of Computer Engineering, Faculty of Computer Science,

Sriwijaya University

Email: meidhihafiz@gmail.com

Abstract

Visualization is a method used to represent data in the form of an image to display hidden information. The visualization in this study uses malware data to be converted into a grayscale image. This study uses 10 types of malware with a total of 1000 data. The test data is divided into training data as much as 80% of the test data is 20% of the total data. Malware is tested using Local Binary Pattern (LBP) to clarify grayscale. The results of classification using K-Nearest Neighbor (K-NN) with values of k = 1, k = 5, k = 10, k = 15, k = 20, k = 25 found an accuracy rate of 96.84%, a precision of 82.01% and F1 score of 81.50%. The results of applying the K-Nearest Neighbor (K-NN) algorithm for malware classification in the form of grayscale images have found very good results.

Keywords : Visualization, *Citra grayscale, Local Binary Pattern, K-Nearest Neighbor*

VISUALISASI DAN KLASIFIKASI *MALWARE* MENGGUNAKAN METODE *K-NEAREST NEIGHBOR*

Meidi Dwi Hafiz (09011281520097)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email: *meidhihafiz@gmail.com*

Abstrak

Visualisasi adalah metode yang digunakan dalam merepresentasikan data kedalam bentuk sebuah gambar untuk menampilkan informasi tersembunyi. Visualisasi pada penelitian ini menggunakan data *malware* untuk di ubah kedalam bentuk *citra grayscale*. Penelitian ini menggunakan 10 jenis *malware* dengan total 1000 data. Data uji dibagi menjadi data latih sebanyak 80% data uji 20% dari total data. Malware diuji menggunakan *Local Binary Pattern (LBP)* untuk mengklarifikasi *grayscale*. Hasil klasifikasi menggunakan *K-Nearest Neighbor* (*K-NN*) dengan nilai k=1, k=5, k=10, k=15, k=20, k=25 mendapati tingkat akurasi sebesar 96.84%, presisi sebesar 82.01% dan F1 score sebesar 81.50%. Hasil penerapan algoritma *K-Nearest Neighbor (K-NN)* untuk klasifikasi *malware* dalam bentuk *citra grayscale* mendapati hasil yang sangat baik.

Kata Kunci : Visualisasi, *Citra grayscale*, *Local Binary Pattern*, *K-Nearest Neighbor*

DAFTAR ISI

DAFTAR GAMBAR

Halaman

DAFTAR TABEL

Halaman

BAB I
PENDAHULUAN

1.1 Pendahuluan

Keamanan ditutun sangat penting karena hari demi hari kejahatan dalam dunia internet terus meningkat. Data pribadi serta informasi penting menjadi sasaran utama bagi pelaku kejahatan. Malware (Malicious Software) merupakan serangkaian intruksi yang bekerja untuk membahayakan suatu sistem. Serangan malware dapat menimbulkan ancaman yang sangat serius terhadap operasi komputer serta informasi penting institusi atau pengguna sehari-hari. Jenis malware pun beragam dengan tujuan yang berbeda seperti Adware, Spyware, Virus, Worm, Trojan, Rootkit, Backdoor, Ransomware [3,7].

Jumlah varian malware serta jenisnya pun meningkat setiap tahun. Alasan utama meningkatnya jumlah varian malware yaitu menghindari deteksi dan penonaktifan oleh anti-virus. Pengembang perangkat lunak berbahaya menggunakan intruksi otomatis untuk menghindari deteksi dengan cara menggunakan kembali malware yang telah dikenali sebelumnya dengan mengubah sebagian kecil code untuk menghasilkan varian baru [2]. Banyak penelitian yang telah dilakukan untuk memecahkan hubungan antar varian malware. Dalam penelitian menggunakan dua pendekatan analisis statis dan dinamis. Analisis statis menganalisis suatu file dengan hanya menemukan struktural penulis dalam menulis kode tanpa menjalankan program [4]. Berbeda dengan analisis statis, dinamis menganalisa biner selama run-time pada lingkungan virtual untuk pemantauan secara langsung perilaku malware [3]. Oleh karena itu, sejumlah besar pekerjaan pada masa ini telah berfokus pada pengembangan alat untuk mengumpulkan, memantau, serta menganalisis malware.

Dalam menganalisis malware yang telah disajikan, para peneliti merepresentasikan malware tersebut kedalam string biner 0 dan 1. File biner yang diberikan dapat dibaca sebagai 8 bit vektor yang kemudian disusun

sebagai 2D array dan setelah itu divisualisasikan menjadi citra grayscale. Gambar berguna untuk mendeteksi perubahan kecil sambil mempertahankan struktur global [6]. Bagian-bagian yang berbeda akan dengan mudah dilihat setelah diekspresikan ke dalam bentuk file gambar grayscale.

Untuk mengatasi volume varian yang besar, Teknik *Machine learning* sangat berguna untuk diterapkan dalam pengklasifikasian malware secara otomatis, Meskipun banyak penelitian yang menerapkannya dengan pencapaian tertentu, akurasi serta efisiensi tetap tidak memadai untuk memenuhi permintaan [5]. Oleh karena itu, kebutuhan akan algoritma yang lebih kuat diperlukan dan menurut penelitian dan penyelidikan yang dilakukan, algoritma pembelajaran mesin adalah algoritma yang ditemukan sangat efisien dan dapat diandalkan. Pada penelitian ini, penulis menerapkan algoritma *K-nearest Neighbor*. Algoritma merupakan suatu metode yang berguna untuk pengklasifikasian terhadap objek berdasarkan dari data pembelajaran dimana diambil jarak terdekat dengan objek tersebut.

1.2 Tujuan

Adapun tujuan dari penelitian ini adalah:

1. Memvisualisaikan Jenis Malware berbeda ke dalam bentuk GrayScale.
2. Menerapkan Algoritma *K-Nearest Neighbor* untuk klasifikasi malware.
3. Menganalisis keakuratan Algoritma *K-Nearest Neighbor* dalam pengklasifikasian malware.

1.3 Manfaat

Adapun manfaat yang dapat diambil dari penelitian ini adalah:

1. Penerapan algoritma *K-Nearest Neighbor* dan pemahaman pada pengklasifikasian malware.
2. Dapat menjadi referensi untuk penelitian selanjutnnya.

1.4 Rumusan dan Batasan Masalah

1. Bagaimana ekstraksi fitur pada citra grayscale malware menggunakan LBP
2. Seberapa tingkat keakurasian sistem *K-Nearest Neighbor* dalam klasifikasi malware yang terlebih dahulu diubah kedalam bentuk citra grayscale.

1.5 Metodologi Penelitian

Penelitian ini menggunakan metodelogi dalam pembuatan tugas akhir yang akan melewati tahapan sebagai berikut:

1. Study Pustaka (Literatur)

   Tahap ini merupakan tahap yang mencari referensi atau literatur pada *Keyword* yang di angkat dari judul yang bertujuan untuk menunjang pada penelitian yang dilakukan.

2. Konsultasi

   Tahap ini, peneliti merupakan tahap dimana penulis melakukan konsultasi rutin kepada orang-orang yang di memiliki pengetahuan serta wawasan terhadap permasalahan yang ditemui pada saat pembuatan Tugas Akhir.

3. Pengumpulan Data

   Tahap ini, data yang didapatkan merupakan data yang berasal dari dataset Microsoft Malware Classification Challenge (BIG 2015), kemudian dipilih beberapa sampel dari tiap malware.

4. Pengolahan Data

   Tahap ini, data yang dilakukan pengolahan data dengan Visualisasi malware kedalam bentuk grayscale dan *K-Nearest Neighbor* untuk tahap klasifikasi.

5. Analisa

   Pada tahap ini, dilakukan pengambilan data dan menganalisa data yang telah dilakukan pengolahan.

6. Kesimpulan dan Saran

   Tahap ini, dilakukan penarikan kesimpulan dari analisa dan studi literatur serta saran untuk penulis selanjutnya jika akan dijadikan bahan referensi.

1.6 Sistematika Penulisan

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

BAB I. PENDAHULUAN

Bab pertama berisi landasan dibuatnya tugas akhir ini.

BAB II. TINJAUAN PUSTAKA

Bab kedua ini terdapat teori dasar tentang tugas akhir ini yaitu *Malware, Machine learning, python, Fitur deskriptor, Local Binary Pattern, K-Nearest Neighbor.*

BAB III. METODOLOGI

Pada bab ketiga merupakan penjelasan sistematis mengenai penelitian yang akan di lakukan, yakni teknik pengumpulan data, pemrosesan data hingga cara kerja dalam pengujuan algoritma.

BAB IV. PENGUJIAN DAN ANALISA

Bab keempat berisikan hasil pengujian yang telah dilakukan, hasil pengujian tersebut akan dianalisis dan akan dilakukan pembuktian sesuai dengan tujuan tugas akhir ini.

BAB V. KESIMPULAN DAN SARAN

Bab kelima merupakan kesimpulan dan saran bagi keseluruhan penelitian 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] L. Zhang and P. N. Suganthan, "Random Forests with ensemble of feature spaces," *Pattern Recognit.*, vol. 47, no. 10, pp. 3429–3437, 2014.

[3] K. V. Uma and E. S. Blessie, "Survey on android malware detection and protection using data mining algorithms," *Proc. Int. Conf. I-SMAC (IoT Soc. Mobile, Anal. Cloud), I-SMAC 2018*, pp. 209–212, 2019.

[4] A. Subasi, S. Alzahrani, A. Aljuhani, and M. Aljedani, "Comparison of Decision Tree Algorithms for Spam E-mail Filtering," *1st Int. Conf. Comput. Appl. Inf. Secur. ICCAIS 2018*, pp. 1–5, 2018.

[5] Y. Li, C. Yan, W. Liu, and M. Li, "A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification," *Appl. Soft Comput. J.*, vol. 70, pp. 1000–1009, 2018.

[6] P. S. Tang, X. L. Tang, Z. Y. Tao, and J. P. Li, "Research on feature selection algorithm based on mutual information and genetic algorithm," *2014 11th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. ICCWAMTIP 2014*, vol. 1, no. 1, pp. 403–406, 2014.

[7] Pfeffer, Avi, et al. "Malware analysis and attribution using genetic information." Malicious and Unwanted Software (MALWARE), 2012 7th International Conference on. IEEE, 2012, pp. 39-45.

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

[9] 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.

[10] Kaggle. (2015). Microsoft Malware Classification Challenge (BIG 2015). Retrieved from https://www.kaggle.com/c/malware-classification

[11] Ojala, T., Pietikäinen, M. dan Maenpaa, T., 2002. Multiresolution Gray Scale dan Rotation Invariant Texture Classification with Local Binary Patterns. *Proceedings - International Conference on Image Processing, ICIP*, 24(7), pp.1852–1855.