

SISTEM KLASIFIKASI ANDROID RANSOMWARE BERBASIS *DEEP NEURAL NETWORK (DNN)*



OLEH:
NURUL AFIFAH
09042611822001

**PROGRAM STUDI MAGISTER ILMU KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
TAHUN 2019**

SISTEM KLASIFIKASI ANDROID RANSOMWARE BERBASIS *DEEP NEURAL NETWORK (DNN)*

TESIS

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



OLEH:

NURUL AFIFAH

09042611822001

**PROGRAM STUDI MAGISTER ILMU KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
TAHUN 2019**

LEMBAR PENGESAHAN
SISTEM KLASIFIKASI ANDROID RANSOMWARE BERBASIS *DEEP NEURAL NETWORK (DNN)*

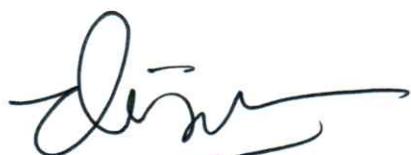
TESIS

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Magister

OLEH :
NURUL AFIFAH
09042611822001

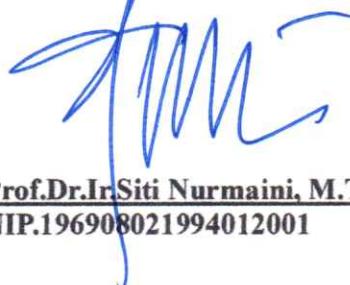
Palembang, November 2019

Pembimbing 1,



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

Pembimbing 2,



Prof. Dr. Ir. Siti Nurmaini, M.T.
NIP.196908021994012001

Mengetahui,
Ketua Program Studi Magister Ilmu Komputer



HALAMAN PERSETUJUAN

Pada hari Jumat tanggal 15 November 2019 telah dilaksanakan ujian sidang tesis oleh Magister Ilmu Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Nurul Afifah

NIM : 09042611822001

Judul : Sistem Klasifikasi Android Ransomware Berbasis Deep Neural Network (DNN)

1. Pembimbing I

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



2. Pembimbing II

Prof. Dr. Ir. Siti Nurmaini, M.T.
NIP. 196908021994012001



3. Pengaji I

Dian Palupi Rini, M.Kom., Ph.D.
NIP. 197802232006042002



4. Pengaji II

Dr. Yusuf Hartono, M.Sc.
NIP. 196411161990031002



Mengetahui,
Ketua Program Studi Magister Ilmu Komputer



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

LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Nurul Afifah
NIM : 09042611822001
Program Studi : Magister Ilmu Komputer
Judul Tesis : Sistem Klasifikasi Android Ransomware Berbasis Deep Neural Network (DNN)

Hasil Pengecekan Software iThenticate/Turnitin : 3 %

Menyatakan bahwa laporan tesis saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan tesis ini, maka saya bersedia menerima sanksi akademik dari universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, November 2019

METERAI
TEMPEL

53D93AHF054860155

6000
ENAM RIBU RUPIAH

(Nurul Afifah)

NIM. 09042611822001

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Allah SWT atas rahmat-nya sehingga penulis dapat menyelesaikan tesis yang berjudul “Sistem Klasifikasi Android Ransomware Berbasis Deep Neural Network (DNN)” di susun untuk memenuhi sebagian persyaratan kelulusan untuk memperoleh gelar Magister Komputer pada Program Studi Magister Ilmu Komputer Universitas Sriwijaya.

Pada kesempatan ini penulis menyadari keterbatasan dan kelemahan yang ada dalam menyelesaikan tesis ini sehingga penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang telah memberikan dukungan, bimbingan dan motivasi kepada penulis untuk menyelesaikan tesis ini, kepada :

1. Bapak Jaidan Jauhari, S.Pd, M.T selaku dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
2. Bapak Dr. Ir. Sukemi, M.T., sebagai Ketua Program Studi Magister Teknik Informatika Universitas Sriwijaya.
3. Bapak Deris Stiawan, M.T., Ph.D dan Prof. Dr. Ir. Siti Nurmaini, M.T selaku pembimbing yang telah meluangkan waktu, bantuan serta saran dan kritiknya dalam penyusunan tesis ini.
4. Terkusus kepada almarhum bapak selaku support system terbesar dalam menyelesaikan s2.
5. Suami tercinta yang selalu mendukung dan terus menyupport dalam penyelesaian s2.
6. Tim COE ajran, ridho yang sudah membantu emak dalam menyelesaikan tesis
7. Seluruh teman-teman Magister Ilmu Komputer yang telah membantu dan memberikan semangat pada masa-masa perkuliahan.

Akhir kata, penulis menyadari bahwa tesis ini masih banyak kekurangan baik dari isi maupun susunan. Semoga tesis ini dapat bermanfaat untuk kita semua.

Palembang, November 2019



Penulis

LEMBAR PERSEMBAHAN

لَا يُكَلِّفُ اللَّهُ نَفْسًا إِلَّا وُسْعَهَا

*“Allah tidak membebani seseorang melainkan sesuai dengan
kesanggupannya”
(QS. Al-Baqarah: 286)*

Kupersembahkan untuk

Almarhum Bapak

Mama

Suamiku dan Anak-Anakku

ANDROID RANSOMWARE CLASSIFICATION SYSTEM BASED ON DEEP NEURAL NETWORK (DNN)

ABSTRACT

Currently there are tons of applications available on the Google Play Store. But some of these applications have been infiltrated by malware files. On Android, each program is contained in apk file, which contains all program code, resources, assets, certificates, and manifest files. Users can download applications from any website and copy the apk file to the device, which further increases the popularity of Android. In open the network still provides many possibilities for the attacker. For example, spread via email and phishing sites. This is the root of the problem if the application that has been infiltrated by the malware file has been installed on the Android user's device, then automatically the file on the Android device can be blocked by malware. The type of malware that is very dangerous and growing rapidly is the type of ransomware. Given their rapid growth, there is an urgent need to develop an effective countermeasure solution by classifying the ransomware using the DNN algorithm and optimizing it by initializing the weights of Xavier and He. The results were significant, with a score of 99.96% for training and testing accuracy, 99.95% for precision, 99.99% for sensitivity, 99.07% for specificity and 99.97% for F1-score. The greater the value obtained from the test results, the better the performance in the testing system.

Keywords: *Ransomware, DNN, Xavier, He, Android*

SISTEM KLASIFIKASI ANDROID RANSOMWARE BERBASIS *DEEP NEURAL NETWORK (DNN)*

ABSTRAK

Saat ini banyak sekali aplikasi yang tersedia di *google play store*, namun tidak menutup kemungkinan beberapa aplikasi tersebut telah disusupi file *malware*. Pada *android*, setiap program terdapat dalam file apk, yang berisi semua kode program, sumber daya, aset, sertifikat, dan file *manifest*. Pengguna dapat mengunduh aplikasi dari situs web apa pun dan menyalin file apk ke perangkat, yang semakin meningkatkan popularitas *android*. Keterbukaan jaringan masih menyediakan banyak kemungkinan bagi penyerang. Misalnya menyebar melalui email dan situs *phishing*. Inilah yang menjadi akar permasalahan jika aplikasi yang telah disusupi file *malware* tersebut sudah diinstal dalam perangkat *android user*, maka otomatis file pada perangkat *android* tersebut bisa di blokir oleh *malware*. Tipe malware yang sangat berbahaya dan berkembang pesat yaitu tipe *ransomware*. Mengingat pertumbuhan mereka yang cepat, ada kebutuhan mendesak untuk mengembangkan solusi penanggulangan yang efektif yaitu dengan mengklasifikasi ransomware menggunakan algoritma DNN dan mengoptimasi dengan inisialisasi bobot Xavier dan He. Hasilnya signifikan yaitu dengan menghasilkan nilai 99.96% untuk akurasi training dan testing, 99.95% untuk presisi, 99.99% untuk sensitivitas, 99.07% untuk spesifisitas dan 99.97% untuk F1-score. Semakin besar nilai yang didapat dari hasil pengujian, semakin bagus performa dalam sistem pengujian.

Kata Kunci: *Ransomware, DNN, Xavier, He, Android*

DAFTAR ISI

| | |
|--|--------------|
| HALAMAN JUDUL | ii |
| HALAMAN PENGESAHAN | iii |
| HALAMAN PERSETUJUAN | iv |
| LEMBAR PERNYATAAN | v |
| KATA PENGANTAR | vi |
| LEMBAR PERSEMPBAHAN | vii |
| ABSTRACT | viii |
| ABSTRAK | ix |
| DAFTAR ISI | x |
| DAFTAR GAMBAR | xiii |
| DAFTAR TABEL | xv |
| DAFTAR SINGKATAN | xvii |
| DAFTAR LAMPIRAN | xviii |
| BAB I PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Perumusan Masalah | 3 |
| 1.3 Batasan Masalah | 4 |
| 1.4 Tujuan | 4 |
| 1.5 Manfaat | 4 |
| 1.6 Metodologi Penulisan | 5 |
| BAB II TINJAUAN PUSTAKA | 6 |
| 2.1 Tinjauan Penelitian | 6 |
| 2.2 Fitur Dataset Malware Ransomware | 8 |
| 2.2.1 <i>Hardware Component</i> | 8 |
| 2.2.2 <i>Requestes Permissions</i> | 8 |
| 2.2.3 <i>Network Adresses</i> | 8 |
| 2.2.4 <i>API Call</i> | 9 |
| 2.3 <i>Autoencoder</i> | 9 |
| 2.4 <i>Deep Neural Network</i> | 10 |
| 2.4.1 Inisialisasi Bobot | 12 |
| 2.4.1.1 Inisialisasi bobot <i>Xavier</i> | 13 |
| 2.4.1.2 Inisialisasi bobot <i>He</i> | 13 |

| | |
|---|----|
| 2.5 <i>Confusion Matrix</i> | 14 |
| BAB III METODE PENELITIAN | 16 |
| 3.1 Studi Pustaka | 16 |
| 3.2 Alur Kerja Metodologi Penelitian | 17 |
| 3.3 Persiapan Data | 19 |
| 3.4 Pengolahan Data Ransomware | 20 |
| 3.4.1 Mempersiapkan Data | 21 |
| 3.4.2 Autoencoder | 21 |
| 3.5 Proses Klasifikasi Menggunakan DNN | 24 |
| 3.5.1 Proses Pelatihan | 26 |
| 3.5.2 Proses Validasi | 27 |
| 3.5.2.1 Validasi <i>Autoencoder</i> | 27 |
| 3.5.2.2 Validasi DNN | 28 |
| 3.5.2.3 Validasi Pembagian data <i>training testing</i> | 28 |
| 3.5.2.4 Validasi Berdasarkan Inisialisasi Bobot | 29 |
| 3.5.3 Proses Pengujian | 29 |
| BAB IV HASIL DAN ANALISIS | 30 |
| 4.1 Hasil Validasi Autoencoder | 30 |
| 4.1.1 Hasil Validasi Autoencoder 2 layer | 30 |
| 4.1.2 Hasil Validasi Autoencoder 3 layer | 32 |
| 4.1.3 Hasil Validasi Autoencoder 4 layer | 35 |
| 4.1.4 Hasil Validasi Autoencoder 5 layer | 37 |
| 4.2 Hasil Validasi Berdasarkan <i>hidden layer</i> DNN | 38 |
| 4.2.1 Hasil Validasi DNN 3 hidden layer | 39 |
| 4.2.2 Hasil Validasi DNN 4 hidden layer | 41 |
| 4.2.3 Hasil Validasi DNN 5 hidden layer | 44 |
| 4.2.4 Hasil Validasi DNN 6 hidden layer | 46 |
| 4.3 Hasil Validasi Pembagian data <i>training testing</i> | 48 |
| 4.3.1 Hasil Validasi data training 60% testing 40% | 48 |
| 4.3.2 Hasil Validasi data training 70% testing 30% | 51 |
| 4.3.3 Hasil Validasi data training 80% testing 20% | 53 |
| 4.3.4 Hasil Validasi data training 90% testing 10% | 55 |

| | |
|---|--------|
| 4.4 Hasil Validasi Berdasarkan Teknik Inisialisasi Bobot | 57 |
| 4.4.1 Hasil Validasi Inisialisasi bobot <i>xavier</i> normal | 57 |
| 4.4.2 Hasil Validasi Inisialisasi bobot <i>xavier</i> uniform | 60 |
| 4.4.3 Hasil Validasi Inisialisasi bobot <i>he</i> normal | 62 |
| 4.4.4 Hasil Validasi Inisialisasi bobot <i>he</i> uniform | 65 |
| 4.5 Analisis Akurasi dan <i>Loss</i> | 67 |
| 4.5.1 Analisis Akurasi dan <i>Loss Autoencoder</i> | 67 |
| 4.5.2 Analisis Akurasi dan <i>Loss</i> DNN | 68 |
| 4.5.3 Analisis Akurasi dan <i>Loss</i> Pembagian Data <i>Training Testing</i> | 69 |
| 4.5.4 Analisis Akurasi dan <i>Loss</i> Berdasarkan Inisialisasi Bobot | 70 |
| 4.6 Analisis Pengujian BACC dan MCC | 71 |
| 4.7 Studi Perbandingan Penelitian | 72 |
| BAB V KESIMPULAN | 75 |
| 5.1 Kesimpulan | 75 |
| 5.2 Saran | 76 |
| DAFTAR PUSTAKA | 77 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2.1 Arsitektur Autoencoder | 9 |
| Gambar 2.2 Arsitektur DNN dengan inisialisasi bobot | 12 |
| Gambar 2.3 <i>Confusion Matrix</i> | 14 |
| Gambar 3.1 Alur Penelitian Keseluruhan | 17 |
| Gambar 3.2 Tahapan Studi Pustaka | 18 |
| Gambar 3.3 Alur Kerja Metodologi Penelitian | 19 |
| Gambar 3.4 Pengolahan Data Ransomware | 20 |
| Gambar 3.5 Sampel dataset Ransomware dan Benign | 21 |
| Gambar 3.6 Arsitektur Autoencoder yang diusulkan | 22 |
| Gambar 3.7 Flowchart Autoencoder | 23 |
| Gambar 3.8 Arsitektur DNN yang diusulkan | 24 |
| Gambar 3.9 Flowchart klasifikasi DNN | 25 |
| Gambar 4.1 Plot Akurasi Autoencoder 2 layer | 30 |
| Gambar 4.2 Plot Loss Autoencoder 2 layer | 31 |
| Gambar 4.3 Plot kurva ROC Autoencoder 2 layer | 32 |
| Gambar 4.4 Plot Akurasi Autoencoder 3 layer | 33 |
| Gambar 4.5 Plot Loss Autoencoder 3 layer | 33 |
| Gambar 4.6 Plot Kurva ROC Autoencoder 3 layer | 34 |
| Gambar 4.7 Plot Akurasi Autoencoder 4 layer | 35 |
| Gambar 4.8 Plot Loss Autoencoder 4 layer | 35 |
| Gambar 4.9 Plot Kurva ROC Autoencoder 4 layer | 36 |
| Gambar 4.10 Plot Akurasi Autoencoder 5 layer | 37 |
| Gambar 4.11 Plot Loss Autoencoder 5 layer | 37 |
| Gambar 4.12 Plot Kurva ROC Autoencoder 5 layer | 38 |
| Gambar 4.13 Plot Akurasi DNN 3 layer | 39 |
| Gambar 4.14 Plot Loss DNN 3 layer | 40 |
| Gambar 4.15 Plot Kurva ROC DNN 3 layer | 41 |
| Gambar 4.16 Plot Akurasi DNN 4 layer | 42 |
| Gambar 4.17 Plot Loss DNN 4 layer | 43 |
| Gambar 4.18 Plot Kurva ROC DNN 4 layer | 43 |
| Gambar 4.19 Plot Akurasi DNN 5 layer | 44 |

| | |
|---|----|
| Gambar 4.20 Plot Loss DNN 5 layer | 44 |
| Gambar 4.21 Plot Kurva ROC DNN 5 layer | 45 |
| Gambar 4.22 Plot Akurasi DNN 6 layer | 46 |
| Gambar 4.23 Plot Loss DNN 6 layer | 47 |
| Gambar 4.24 Plot Kurva ROC DNN 6 layer | 48 |
| Gambar 4.25 Plot Akurasi data training 60% dan testing 40% | 49 |
| Gambar 4.26 Plot Loss data training 60% dan testing 40% | 49 |
| Gambar 4.27 Plot Kurva ROC data training 60% dan testing 40% | 50 |
| Gambar 4.28 Plot Akurasi data training 70% dan testing 30% | 51 |
| Gambar 4.29 Plot Loss data training 70% dan testing 30% | 51 |
| Gambar 4.30 Plot Kurva ROC data training 70% dan testing 30% | 52 |
| Gambar 4.31 Plot Akurasi data training 80% dan testing 20% | 53 |
| Gambar 4.32 Plot Loss data training 80% dan testing 20% | 53 |
| Gambar 4.33 Plot Kurva ROC data training 80% dan testing 20% | 54 |
| Gambar 4.34 Plot Akurasi data training 90% dan testing 10% | 55 |
| Gambar 4.35 Plot Loss data training 90% dan testing 10% | 56 |
| Gambar 4.36 Plot Kurva ROC data training 90% dan testing 10% | 57 |
| Gambar 4.37 Plot Akurasi Inisialisasi bobot Xavier normal | 58 |
| Gambar 4.38 Plot Loss Inisialisasi bobot Xavier normal | 58 |
| Gambar 4.39 Plot Kurva ROC Inisialisasi bobot Xavier normal | 59 |
| Gambar 4.40 Plot Akurasi Inisialisasi bobot Xavier uniform | 60 |
| Gambar 4.41 Plot Loss Inisialisasi bobot Xavier uniform | 60 |
| Gambar 4.42 Plot Kurva ROC Inisialisasi bobot Xavier uniform | 61 |
| Gambar 4.43 Plot Akurasi Inisialisasi bobot He normal | 62 |
| Gambar 4.44 Plot Loss Inisialisasi bobot He normal | 63 |
| Gambar 4.45 Plot Kurva ROC Inisialisasi bobot He normal | 64 |
| Gambar 4.46 Plot Akurasi Inisialisasi bobot He uniform | 65 |
| Gambar 4.47 Plot Loss Inisialisasi bobot He uniform | 65 |
| Gambar 4.48 Plot Kurva ROC Inisialisasi bobot He uniform | 66 |
| Gambar 4.49 Analisis Plot Akurasi dan <i>Loss Autoencoder</i> | 67 |
| Gambar 4.50 Analisis Plot Akurasi dan <i>Loss DNN</i> | 68 |
| Gambar 4.51 Analisis Plot Akurasi dan <i>Loss Pembagian Data</i> | 69 |
| Gambar 4.52 Analisis Plot Akurasi dan <i>Loss Inisialisasi Bobot</i> | 70 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2.1 Penelitian tentang malware 5 tahun terakhir | 7 |
| Tabel 2.2 Perbedaan Inisialisasi bobot Xavier normal dan uniform | 13 |
| Tabel 2.3 Perbedaan Inisialisasi bobot He normal dan uniform | 13 |
| Tabel 3.1 Jumlah Data Ransomware Benign | 20 |
| Tabel 3.2 Pembagian Data Pelatihan dan Validasi | 27 |
| Tabel 3.3 Struktur dan Hyperparameter Autoencoder | 27 |
| Tabel 3.4 Struktur dan Hyperparameter dari Validasi DNN | 28 |
| Tabel 3.5 Struktur dan Hyperparameter dari pembagian data | 28 |
| Tabel 3.6 Struktur dan Hyperparameter dari Inisialisasi bobot | 29 |
| Tabel 4.1 Hasil Performa Validasi Autoencoder 2 layer | 31 |
| Tabel 4.2 Nilai Confusion Matrix Autoencoder 2 layer | 32 |
| Tabel 4.3 Hasil Performa Validasi Autoencoder 3 layer | 34 |
| Tabel 4.4 Nilai Confusion Matrix Autoencoder 3 layer | 34 |
| Tabel 4.5 Hasil Performa Validasi Autoencoder 4 layer | 36 |
| Tabel 4.6 Nilai Confusion Matrix Autoencoder 4 layer | 37 |
| Tabel 4.7 Hasil Performa Validasi Autoencoder 5 layer | 38 |
| Tabel 4.8 Nilai Confusion Matrix Autoencoder 5 layer | 39 |
| Tabel 4.9 Hasil Performa Validasi DNN 3 layer | 40 |
| Tabel 4.10 Nilai Confusion Matrix DNN 3 layer | 41 |
| Tabel 4.11 Hasil Performa Validasi DNN 4 layer | 43 |
| Tabel 4.12 Nilai Confusion Matrix DNN 4 layer | 44 |
| Tabel 4.13 Hasil Performa Validasi DNN 5 layer | 45 |
| Tabel 4.14 Nilai Confusion Matrix DNN 5 layer | 46 |
| Tabel 4.15 Hasil Performa Validasi DNN 6 layer | 47 |
| Tabel 4.16 Nilai Confusion Matrix DNN 6 layer | 48 |
| Tabel 4.17 Hasil Performa Validasi Data training 60% testing 40% | 49 |
| Tabel 4.18 Nilai Confusion Matrix Data training 60% testing 40% | 50 |
| Tabel 4.19 Hasil Performa Validasi Data training 70% testing 30% | 52 |
| Tabel 4.20 Nilai Confusion Matrix Data training 70% testing 30% | 53 |
| Tabel 4.21 Hasil Performa Validasi Data training 80% testing 20% | 54 |
| Tabel 4.22 Nilai Confusion Matrix Data training 80% testing 20% | 55 |

| | |
|--|----|
| Tabel 4.23 Hasil Performa Validasi Data training 90% testing 10% | 56 |
| Tabel 4.24 Nilai Confusion Matrix Data training 90% testing 10% | 57 |
| Tabel 4.25 Hasil Performa Validasi Inisialisasi bobot Xavier normal | 59 |
| Tabel 4.26 Nilai Confusion Matrix Inisialisasi bobot Xavier normal | 60 |
| Tabel 4.27 Hasil Performa Validasi Inisialisasi bobot Xavier uniform | 61 |
| Tabel 4.28 Nilai Confusion Matrix Inisialisasi bobot Xavier uniform | 62 |
| Tabel 4.29 Hasil Performa Validasi Inisialisasi bobot He normal | 63 |
| Tabel 4.30 Nilai Confusion Matrix Inisialisasi bobot He normal | 64 |
| Tabel 4.31 Hasil Performa Validasi Inisialisasi bobot He uniform | 66 |
| Tabel 4.32 Nilai Confusion Matrix Inisialisasi bobot He unifom | 67 |
| Tabel 4.33 Hasil Performa Evaluasi dalam BACC dan MCC Pengujian | 68 |
| Tabel 4.34 Perbandingan Penelitian Klasifikasi Malware 5 tahun terakhir | 69 |

DAFTAR SINGKATAN

| | |
|------|-------------------------------------|
| DNN | = Deep Neural Network |
| TP | = True Positif |
| TN | = True Negatif |
| FP | = False Positif |
| FN | = False Negatif |
| GN | = Girvan Newman |
| DT | = Decision Tree |
| SVM | = Support Vector Machine |
| CNN | = Convolutional Neural Network |
| DBN | = Deep Belief Network |
| BACC | = Balanced Accuracy |
| MCC | = Matthew's Correlation Coefficient |

DAFTAR LAMPIRAN

Lampiran 1 Source Code Python

Lampiran 2 Jurnal

BAB I. PENDAHULUAN

Pendahuluan bab ini menjelaskan tentang latar belakang penelitian yang berjudul: “Sistem Klasifikasi *Android Ransomware* Berbasis *Deep Neural Network*”. Permasalahan *malware ransomware* dipilih menjadi topik dikarenakan jenis serangan ini yang paling berbahaya jika sudah menyerang perangkat dan mengakibatkan perangkat tersebut tidak bisa dibuka dan *hacker* meminta tebusan jika ingin membuka akses file tersebut Chen dkk. (2018). Pengimplementasian *Deep Learning* sebagai algoritma klasifikasi agar fitur dipelajari secara otomatis. *Deep Neural Network* (DNN) adalah algoritma yang digunakan dalam penelitian ini dikarenakan karakter data yang bersifat vektor tabular.

1.1 Latar Belakang

Malware merupakan file berbahaya yang dapat menyusup ke sistem operasi sehingga dapat merusak sistem dan juga dapat mencuri file penting pada sistem. Pada publikasi Idika dan Mathur (2007) *malware* mencakup virus komputer, *trojan horse*, *ransomware* dan perangkat lunak lainnya yang berniat jahat dan tidak diinginkan. Seiring berkembangnya teknologi pun memiscu dikembangkannya *malware* baru, sehingga semakin banyak pihak yang dirugikan karena adanya *malware* ini. Bahkan saat ini, *malware* telah menjangkit hampir seluruh jenis sistem operasi salah satunya sistem operasi *android*. Sistem *android* memiliki kelebihan dibanding sistem operasi lain yakni bersifat *open access*, Akan tetapi, salah satu keunggulan sistem *android* menjadi salah satu kelemahannya. Bersifat *open access*, dimana *user* dapat menciptakan dan mengembangkan aplikasi sendiri sehingga dapat digunakan pada bermacam perangkat seluler. Hal ini malah menimbulkan kemudahan pada pihak yang tidak bertanggung jawab untuk membangun dan mengembangkan *malware* menjadi aplikasi yang dapat masuk ke system Arp dkk. (2014).

Saat ini banyak sekali aplikasi yang tersedia di *google play store*, namun tidak menutup kemungkinan beberapa aplikasi tersebut telah disusupi file *malware*. Pada *android*, setiap program terdapat dalam file apk, yang berisi semua kode program, sumber daya, asset, sertifikat, dan file *manifest*.

Pengguna dapat mengunduh aplikasi dari situs web apa pun dan menyalin file apk ke perangkat, yang semakin meningkatkan popularitas *android*. Sayangnya, popularitas seperti itu juga menarik perhatian peneliti *malware*. Bahkan jika pasar resmi dan pihak ketiga menyebarkan skema keamanan untuk mencegah penyebaran *malware*, keterbukaan jaringan masih menyediakan banyak kemungkinan bagi penyerang. Misalnya menyebar melalui email dan situs *phishing* Arp dkk. (2014). Inilah yang menjadi akar permasalahan jika aplikasi yang telah disusupi file *malware* tersebut sudah diinstal dalam perangkat *android user*, maka otomatis file pada perangkat *android* tersebut bisa diblokir oleh *malware*. Didalam publikasi Chen dkk. (2018) ditemukan *malware* yang dapat memblokir data pada perangkat *android* yaitu *malware* tipe *ransomware*.

Tingkat infeksi paling banyak oleh *ransomware android* adalah di Eropa Timur dan Rusia hingga 59%, diikuti oleh Inggris dan AS (25%), dan kemudian Cina (16%). Mengingat pertumbuhan mereka yang cepat, ada kebutuhan mendesak untuk mengembangkan solusi penanggulangan yang efektif Chen dkk. (2018)

Algoritma konvensional telah diusulkan untuk mengklasifikasi *malware* yaitu Algoritma *Girvan-Newman* (GN). Algoritma tersebut melatih klasifikasi biner. Metode *machine learning* seperti yang dilakukan oleh Lashkari dkk. (2018) menyatakan bahwa dengan menggunakan metode *Decision Tree* (DT) dapat menghasilkan *Precision* 85% dan *Recall* 88% dalam mengklasifikasi jenis *malware*. Penelitian Dan Lo dkk. (2016) menghasilkan akurasi 99,60%. Penelitian Yerima dan Sezer (2019) menghasilkan tingkat sensitivitas sebesar 95,80% dalam mengklasifikasi *malware*. Metode *machine learning* seperti yang dilakukan oleh Lashkari dkk. (2018), Dan Lo dkk. (2016), Yerima dan Sezer (2019) menggunakan metode SVM, *Decision tree* dan *Random forest*. Metode *machine learning* tersebut masih mempelajari fitur secara manual (ekstraksi fitur) sehingga muncul paradigma *deep learning*. Contoh penerapan teknik *deep learning* dalam permasalahan *malware* antara lain *Deep Neural Network* (DNN), *Convolutional Neural Network*

(CNN) dan *Deep Belief Network* (DBN). Metode yang diusulkan dalam penyelesaian tesis ini adalah DNN. Hal ini dikarenakan karakteristik dari dataset *malware* tersebut yang bersifat matriks vector tabulasi dan *cross section* sangat cocok jika diimplementasikan menggunakan metode DNN (Zhou, 2018). DNN mampu menghasilkan kinerja yang lebih baik dan lebih unggul dalam hal klasifikasi dengan cara mempelajari fitur secara otomatis. Teknik dalam pembelajaran fitur otomatis yaitu *autoencoder* tercantum dalam penelitian Naway dan Li (2019) dan Le dkk. (2018). *Autoencoder* merupakan model pembelajaran fitur otomatis dalam DNN yang mampu melakukan *dimensionality reduction* terhadap data tanpa mengurangi informasi yang ada dalam dataset tersebut Kunang dkk. (2019). Namun implementasi DNN pada kasus klasifikasi *malware* pada penelitian Cui dkk. (2018) memiliki kelemahan yaitu pada saat *training* data terjadi ketidakstabilan nilai *gradien*. Ketidakstabilan nilai *gradien* dapat mempengaruhi tingkat *konvergen* dan waktu dalam *training* data. Itu dapat ditanggulangi dengan menerapkan teknik inisialisasi bobot dari penelitian Kwon dkk. (2019) dan He dkk. (2015). Penelitian ini dianggap sangat berguna agar dapat mengklasifikasi *android ransomware* dengan hasil klasifikasi yang lebih *konvergen* antara data *training* dan *testing*.

1.2 Perumusan Masalah

DNN merupakan algoritma yang memiliki performa sangat baik untuk klasifikasi malware pada penelitian Afifah dkk. (2019) karena dapat menyelesaikan permasalahan klasifikasi *malware*. DNN menginisialisasi bobot secara acak dalam proses klasifikasi. Apabila nilai bobot awal terlalu kecil maka *input* ke setiap lapisan *hidden layer* dan *output* akan sangat kecil yang dapat menyebabkan terjadinya *vanishing gradient*. Jika bobot awal terlalu besar maka terjadi *exploding gradient* seperti pada pada penelitian Karbab dkk. (2018) yang memicu nilai *loss* menjadi naik turun pada proses pelatihan data sehingga menjadi tidak konvergen berdasarkan penelitian Kwon dkk. (2019). Berdasarkan latar belakang diatas, terdapat beberapa masalah yang dirumuskan dalam penelitian ini yaitu:

1. Bagaimana cara mengatasi ketidakstabilan nilai *gradien* dalam proses klasifikasi DNN?
2. Bagaimana cara menerapkan inisialisasi bobot *Xavier* dan *He*?

3. Bagaimana pengaruh inisialisasi bobot terhadap nilai akurasi, sensitivitas, spesifisitas, presisi dan *F1 score* terhadap kinerja dari sistem klasifikasi *android ransomware* menggunakan DNN?

1.3 Batasan Masalah

Terdapat beberapa batasan masalah yang dirancang dalam tesis ini yaitu:

1. Data yang digunakan merupakan dataset dari *Drebin Project* Arp dkk. (2014).
2. Analisis *malware ransomware* dan *benign* menggunakan *static analysis* Arp dkk. (2014) Naway dan Li (2018)
3. Proses *dimentinality reduction* yang digunakan adalah *Autoencoder*
4. Inisialisasi bobot menggunakan Persamaan *Xavier* dan *He*.

1.4 Tujuan

Tujuan dalam penelitian tesis ini adalah sebagai berikut :

1. Mengetahui teknik inisialisasi bobot terbaik yang dapat menyelesaikan permasalahan ketidakstabilan *gradien*
2. Menerapkan teknik inisialisasi bobot *Xavier* dan *He* sehingga didapat grafik yang konvergen dari nilai akurasi dan *loss* dari proses *training* data
3. Mengukur akurasi, sensitivitas, spesifisitas, presisi dan *F1 score* terhadap kinerja pengklasifikasi DNN

1.5 Manfaat

Hasil dari penelitian ini dapat menjadi landasan dalam pengembangan sistem klasifikasi *android ransomware* secara lebih lanjut. Selain itu manfaat dari penelitian ini secara praktis yaitu sebagai berikut:

1. Inisialisasi bobot menggunakan persamaan *Xavier* dan *He* berhasil membuat hasil dari proses *training* data menjadi konvergen
2. Hasil dari penelitian ini dapat menjadi referensi untuk meningkatkan nilai performa akurasi, sensitivitas, spesifisitas, presisi dan *F1 score* dalam sistem klasifikasi yang menerapkan metode DNN

1.6 Metodologi Penulisan

Agar memperoleh gambaran jelas mengenai penelitian ini, maka dibuatlah suatu sistematika penulisan yang berisi gambaran dalam tiap bab penelitian ini, yaitu:

- | | |
|-------------------|--|
| 1. BAB I | Pendahuluan |
| | Bab ini menjelaskan tentang latar belakang, perumusan masalah, batasan masalah, tujuan dan manfaat dari topik yang dipilih berupa sistem klasifikasi data file normal dan serangan <i>malware ransomware</i> menggunakan metode DNN. |
| 2. BAB II | Tinjauan Pustaka |
| | Bab ini menjelaskan mengenai <i>literature review</i> yang berhubungan dengan masalah <i>malware</i> dengan metode DNN yang mengacu pada beberapa penelitian publikasi. |
| 3. BAB III | Metodologi Penelitian |
| | Bab ini menjelaskan pembahasan secara bertahap dan rinci langkah yang digunakan untuk mengumpulkan dan menganalisa jenis file normal dan jenis file <i>malware</i> . Metodologi ini menjelaskan pendekatan algoritma DNN serta model yang digunakan sehingga tujuan dari penulisan tercapai. |
| 4. BAB IV | Analisa dan Pembahasan |
| | Bab ini berisi hasil pengujian yang dilakukan, data yang diuji akan dianalisa menggunakan berbagai macam teknik serta validasi hasil. |
| 5. BAB V | Kesimpulan |
| | Bab ini menjelaskan kesimpulan dari hasil yang diperoleh, serta merupakan jawaban yang diperoleh dari tujuan yang ingin dicapai. |

DAFTAR PUSTAKA

- Afifah, N., Stiawan, D., & Nurmaini, S. (2019). The Implementation of Deep Neural Networks Algorithm for Malware Classification. *Computer Engineering Applications Journal*8(3), 189–202.
- Arp, D., Spreitzenbarth, M., Malte, H., Gascon, H., & Rieck, K. (2013). DREBIN: Effective and Explainable Detection of Android Malware in Your Pocket Daniel. *Choice Reviews Online*, 45(02), 45-0765-45-0765. <https://doi.org/10.5860/choice.45-0765>
- Brodersen, K. H., Soon Ong, C., Stephan, K. E., & Buhmann, J. M. (2010). *The balanced accuracy and its posterior distribution*. <https://doi.org/10.1109/ICPR.2010.764>
- Chen, J., Wang, C., Zhao, Z., Chen, K., Du, R., & Ahn, G. J. (2018). Uncovering the Face of Android Ransomware: Characterization and Real-Time Detection. *IEEE Transactions on Information Forensics and Security*, 13(5), 1286–1300. <https://doi.org/10.1109/TIFS.2017.2787905>
- Cui, Z., Xue, F., Cai, X., Cao, Y., Wang, G., Chen, J., & Member, S. (2018). *Detection of Malicious Code Variants Based on Deep Learning*. <https://doi.org/10.1109/TII.2018.2822680>
- Dan Lo, C. T., Ordóñez, P., & Cepeda, C. (2016). Feature selection and improving classification performance for malware detection. *Proceedings - 2016 IEEE International Conferences on Big Data and Cloud Computing, BDCloud 2016, Social Computing and Networking, SocialCom 2016 and Sustainable Computing and Communications, SustainCom 2016*, 560–566. <https://doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.87>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE International Conference on Computer Vision, 2015 Inter*, 1026–1034. <https://doi.org/10.1109/ICCV.2015.123>
- Idika, N., & Mathur, A. P. (2007). *A Survey of Malware Detection Techniques. initialization weight malware*. (n.d.).

- Kelimeler, A. (n.d.). *Zararlı Mobil Uygulamaların Derin Yapay Sinir Ağı ile Tespit Edilmesi Mobile Malware Detection Using Deep Neural Network.*
- Kunang, Y. N., Nurmaini, S., Stiawan, D., Zarkasi, A., & Jasmir, F. (2019). Automatic Features Extraction Using Autoencoder in Intrusion Detection System. *Proceedings of 2018 International Conference on Electrical Engineering and Computer Science, ICECOS 2018*, (June 2019), 219–224. <https://doi.org/10.1109/ICECOS.2018.8605181>
- Lashkari, A. H., Kadir, A. F. A., Taheri, L., & Ghorbani, A. A. (2018). Toward Developing a Systematic Approach to Generate Benchmark Android Malware Datasets and Classification. *Proceedings - International Carnahan Conference on Security Technology, 2018-Octob(Cic)*, 1–7. <https://doi.org/10.1109/CCST.2018.8585560>
- Le, Q., Boydell, O., Namee, B. Mac, & Scanlon, M. (2018). Deep learning at the shallow end: Malware classification for non-domain experts. *Digital Investigation*, 26, S118–S126. <https://doi.org/10.1016/j.diin.2018.04.024>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Luque, A., Carrasco, A., Martín, A., & De Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91, 216–231. <https://doi.org/10.1016/j.patcog.2019.02.023>
- MacLennan, B. J. (2016). Field computation: A framework for quantum-inspired computing. In *Quantum Inspired Computational Intelligence: Research and Applications*. <https://doi.org/10.1016/B978-0-12-804409-4.00003-6>
- Naway, A., & Li, Y. (2018). International Journal of Computer Science and Mobile Computing A Review on The Use of Deep Learning in Android Malware Detection. In *International Journal of Computer Science and Mobile Computing* (Vol. 7).
- Naway, A., & Li, Y. (2019). *Android Malware Detection Using Autoencoder*. 1–9. Retrieved from <http://arxiv.org/abs/1901.07315>
- Nix, R., & Zhang, J. (n.d.). *Classification of Android Apps and Malware Using Deep Neural Networks.*

- Nurmaini, S., Umi P, R., Naufal, M., & Gani, A. (2018). Cardiac Arrhythmias Classification Using Deep Neural Networks and Principle Component Analysis Algorithm. In *Int. J. Advance Soft Compu. Appl* (Vol. 10).
- Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). *Activation Functions: Comparison of trends in Practice and Research for Deep Learning*. 1–20. Retrieved from <http://arxiv.org/abs/1811.03378>
- Xavier Glorot, Y. B. (1993). Initializing weights to a hidden layer of a multilayer neural network by linear programming. *Proceedings of the International Joint Conference on Neural Networks*, 2, 1701–1704.
- Yerima, S. Y., & Sezer, S. (2019). DroidFusion: A Novel Multilevel Classifier Fusion Approach for Android Malware Detection. *IEEE Transactions on Cybernetics*, 49(2), 453–466. <https://doi.org/10.1109/TCYB.2017.2777960>
- Zhou, B. (2018). Deep Learning and the Cross-Section of Stock Returns: Neural Networks Combining Price and Fundamental Information. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3179281>