

**IMPLEMENTASI ALGORITMA PCA UNTUK PENINGKATAN  
SISTEM *CLASSIFICATION MALWARE ANDROID***

**TUGAS AKHIR**



**OLEH :  
MUHAMMAD NAWAWI  
09011381722104**

**JURUSAN SISTEM KOMPUTER  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS SRIWIJAYA  
2021**

**LEMBAR PENGESAHAN**

**IMPLEMENTASI ALGORITMA PCA UNTUK PENINGKATAN SISTEM  
KLASIFIKASI *MALWARE ANDROID***

**PROPOSAL TUGAS AKHIR**

**Program Studi Sistem Komputer  
Jenjang S1**

**Oleh**

**Muhammad Nawawi  
09011381722104**

**Palembang,**


**Mengetahui,**

**Ketua Jurusan Sistem Komputer**



**Dr. Ir. H. Sakemi, M.T**  
**NIP. 196612032006041001**

**Pembimbing Tugas Akhir**



**Deris Stiawan, Ph. D.**  
**NIP. 197806172006041002**

**HALAMAN PERSETUJUAN**

Telah diuji dan lulus pada:

Hari : Kamis

Tanggal : 01 April 2021

Tim Penguji:

1. Ketua : Sarmayanta Sembiring, M.T.
2. Sekretaris : Sri Desy Siswanti, M.T.
3. Anggota : Huda Ubaya, M.T.
4. Pembimbing : Deris Stiawan, Ph.D.



By date of 2.55.51, 00/00/21

Mengetahui,

Ketua Jurusan Sistem Komputer



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

NIP. 196612032006041001

**SURAT KETERANGAN PENGECEKKAN  
SIMILIARITY**

Saya yang bertanda tangan dibawah ini :

Nama : Muhammad Nawawi  
NIM : 09011381722104  
Prodi : Sistem Komputer  
Fakultas : Ilmu Komputer

Menyatakan bahwa benar hasil pengecekan similarity Skripsi/Tesis/Disertasi/ Lap. Penelitian yang berjudul IMPLEMENTASI ALGORITMA PCA UNTUK PENINGKATAN SISTEM CLASSIFICATION MALWARE ANDROID adalah 19%.

Dicek oleh operator \*:

1. Dosen Pembimbing
- ② UPT Perpustakaan
3. Operator Fakultas .....

Demikian surat keterangan ini saya buat dengan sebenarnya dan dapat saya pertanggung jawabkan.

Indralaya, Mei 2021

Menyetujui

Dosen pembimbing



Deris Stiawan, Ph.D.  
Nip. 197806172006041002

Yang menyatakan



Muhammad Nawawi  
Nim. 09011381722104

\*Lingkari salah satu jawaban tempat anda melakukan pengecekan Similarity

## **MOTTO :**

**Saat semua orang menganggapmu tidak bisa apa-apa.**

**Jangan perdulikan,**

**Karena**

**Yang bisa mengubah nasibmu,**

**Adalah kamu,**

**Bukan orang lain.**

**Dan**

**Ketika dunia jahat kepadamu,**

**Maka berusahalah untuk menghadapinya,**

**Karena tidak akan ada orang yang membantumu**

**Jika kau tidak berusaha.**

**Ku persembahkan untuk:**

- **Papa, Mama, Adik-Adik, Kakak, Nenek, Kakek, Makwo, Pakwo, Bibik, Oom, Encik, Adinda yang selalu mendukung dan memberikan semangat kepadaku.**
- **Teman-teman seperjuangan Sistem Komputer Universitas Sriwijaya angkatan 2017 yang tidak akan kulupakan**
  - **Almamaterku Universitas Sriwijaya**

## KATA PENGANTAR



Puji dan syukur penulis panjatkan kehadirat Allah SWT, atas segala karunia dan rahmat-Nya sehingga penulis dapat menyelesaikan proposal tugas akhir dengan judul **“Implementasi Algoritma PCA Untuk Peningkatan Sistem *Classification Malware Android*”**.

Penulisan Proposal Tugas Akhir ini dilakukan untuk melengkapi salah satu syarat memperoleh gelar Sarjana Komputer di Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya. Adapun sebagai bahan penulisan, penulis mengambil berdasarkan hasil penelitian, observasi dan beberapa sumber literatur yang mendukung dalam penulisan proposal ini. Pada kesempatan ini juga, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu baik dari segi moril ataupun materil serta memberikan kemudahan, dorongan, saran dan kritik selama dalam proses penulisan Proposal Tugas Akhir ini. Oleh karena itu, pada kesempatan ini penulis mengucapkan rasa syukur kepada Allah SWT. dan mengucapkan terima kasih kepada yang terhormat :

1. Ke empat Orang Tua terutama untuk mama yang slalu mensupport dan mendoakan serta keluarga penulis tercinta, yang telah memberikan doa dan restu serta dukungan yang sangat besar selama mengikuti dan melaksanakan perkuliahan di Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya hingga dapat menyelesaikan Proposal Tugas Akhir ini.
2. Bapak Jaidan Jauhari, S.Pd., M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak DR. IR. H. Sukemi, M. T selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Deris Stiawan, Ph. D. selaku Dosen Pembimbing Tugas Akhir penulis
5. Bapak Rossi Passarella, M.ENG selaku Dosen Akademik di Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
6. dosen, staff, serta karyawan Fakultas Ilmu Komputer Universitas Sriwijaya.

7. Seluruh teman-teman seperjuangan angkatan 2017 Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya terutama Abdi, Vanisa, Ika, Adinda, Barzan, Tata, Marle, Hafidz, Hadi, Naufal, Yuan, Badruz, Tri Agung, Taufik, Ryan, Tiara, Nanda, Vira dan Fidy.
8. Almamater.

Penulis menyadari bahwa masih banyak kekurangan dalam penulisan Proposal Tugas Akhir ini. Karena sesungguhnya tak ada yang sempurna didunia ini. Untuk itu, segala saran dan kritik sangatlah penting bagi penulis. Akhir kata, semoga Proposal Tugas Akhir ini dapat bermanfaat dan berguna bagi khalayak.

Palembang, April 2021  
Penulis

**Muhammad Nawawi**  
**NIM. 09011381722104**

**IMPLEMENTATION OF PCA ALGORITHM TO IMPROVE ANDROID  
MALWARE CLASSIFICATION SYSTEM**

**Muhammad Nawawi (09011381722104)**

*Department of Computer Engineering, Faculty of Computer Science,  
Sriwijaya University*

Email : muhammadnawawi1516@gmail.com

***Abstract***

*Dimensional reduction has been widely used in various studies in the world because of its good function in reducing data without removing the characteristics of the data. In this study, we want to see how the increase in results will be obtained for classifying if you use dimensional reduction in the data you want to use. In this research, the dimensional reduction of the PCA algorithm is used to classify android malware, namely adware. The classification in this study was carried out using three classification using ANN (Artificial Neural Network). Differentiate each component from the result of dimensionality reduction using PCA and using optimal functions such as Adam, Adadelta and SGD (Stochastic Gradient Descent). Get pretty good results with a value of 99.90% accuracy, 99.99% precision, 99.84% sensitivity, 99.98% specificity and 99.91% for the F1-Score.*

***Keywords*** : *Malware Android, Adware, Reduksi dimensi, algoritma PCA, Classification, Artificial Neural Network.*



# IMPLEMENTASI ALGORITMA PCA UNTUK PENINGKATAN SISTEM *CLASSIFICATION MALWARE ANDROID*

**Muhammad Nawawi (09011381722104)**

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email : muhammadnawawi1516@gmail.com

## **Abstrak**

Reduksi dimensi sudah banyak dipakai berbagai penelitian didunia karena fungsinya yang baik dalam memperkecil data tanpa menghilangkan karakteristik dari data tersebut. Pada penelitian ini ingin melihat bagaimana peningkatan hasil yang akan didapatkan untuk melakukan klasifikasi jika menggunakan reduksi dimensi pada data yang ingin digunakan. Pada penelitian kali ini menggunakan reduksi dimensi algoritma PCA untuk melakukan klasifikasi pada malware android yaitu adware. Klasifikasi pada penelitian ini dilakukan dengan menggunakan tiga buah *classification* dengan menggunakan ANN (*Artificial Neural Network*). Membedakan tiap komponen dari hasil reduksi dimensi menggunakan PCA dan menggunakan fungsi optimal seperti Adam, Adadelta dan SGD (*Stochastic Gradient Descent*). Mendapatkan hasil yang cukup baik yaitu nilai 99.90% akurasi, 99.99% presisi, 99.84% sensitivitas, 99.98% spesifitas dan 99.91% untuk F1-Score.

**Kata Kunci :** *Malware Andorid, Adware, Reduksi dimensi, algoritma PCA, Klasifikasi, Artificial Neural Network.*

## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	i
<b>HALAMAN PENGESAHAN</b> .....	ii
<b>HALAMAN PERSETUJUAN</b> .....	iii
<b>HALAMAN PERNYATAAN</b> .....	iv
<b>MOTTO</b> .....	v
<b>KATA PENGANTAR</b> .....	vi
<i>Abstract</i> .....	viii
<b>Abstrak</b> .....	ix
<b>DAFTAR ISI</b> .....	x
<b>DAFTAR GAMBAR</b> .....	xiii
<b>DAFTAR TABEL</b> .....	xv
<b>DAFTAR LAMPIRAN</b> .....	xvii
<b>BAB I PENDAHULUAN</b>	
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan.....	3
1.5 Manfaat.....	3
1.6 Metodologi Penelitian .....	4
1.7 Sistematika Penulisan.....	4
<b>BAB II TINJAUAN PUSTAKA</b>	
2.1 Tinjauan Penelitian.....	6
2.2 <i>Android</i> .....	8
2.3 <i>Malware</i> .....	8
2.4 Dataset CICAndMal2017 .....	9
2.5 <i>Dimentional Reduction</i> .....	10
2.5.1 <i>AutoEncoder (AE)</i> .....	11
2.5.2 <i>Principal Component Analysis (PCA)</i> .....	11
2.6 <i>Artificial Neural Network (ANN)</i> .....	12
2.6.1 <i>Komponen Artificial Neural Network (ANN)</i> .....	13
2.6.2 <i>Arsitektur Artificial Neural Network (ANN)</i> .....	13

2.6.3 Fungsi Aktvasi .....	14
2.6.3.1 <i>Linear Activation Function</i> .....	14
2.6.3.2 <i>Non-Linear Activation Function</i> .....	14
2.6.3.2.1 <i>Sigmoid</i> atau <i>Logistic Activation Function</i> .....	15
2.6.3.2.2 <i>Tanh</i> atau <i>Hyperbolic Tangent Activation Function</i> .....	15
2.6.3.2.3 <i>ReLU (Retified Linear Unit) Activation Function</i> .....	16
2.6.4 Fungsi Pengoptimal .....	17
2.6.4.1 <i>Adam</i> .....	17
2.6.4.2 <i>Adadelata</i> .....	17
2.6.4.3 <i>Stochastic Gradient Descent (SGD)</i> .....	18
2.6.5 Fungsi <i>Loss</i> .....	18
2.6.5.1 <i>Binary</i> .....	18
2.6.5.2 <i>Multi-Class</i> .....	19
2.7 <i>Confusion Matrix</i> .....	19
2.8 <i>CICFlowMeter</i> .....	20
2.9 <i>Python</i> .....	21
2.9.1 <i>Jupyter Notebook</i> .....	21

### **BAB III METODOLOGI**

3.1 Raw Data .....	23
3.2 Ekstraksi PCAP .....	24
3.3 <i>Pre-prossesing</i> Data .....	26
3.4 Reduksi Dimensi.....	28
3.5 Klasifikasi.....	30
3.5.1 Proses Pelatihan .....	35
3.5.2 Proses Validasi.....	36
3.5.2.1 Validasi PCA .....	36
3.5.2.2 Validasi ANN .....	37
3.6 Evaluasi Model .....	37
3.6.1 Peningkatan Sistem Classification.....	39

3.6.2 Perbandingan Classification menggunakan PCA dan tanpa PCA	39
-----------------------------------------------------------------	----

## **BAB IV HASIL DAN PEMBAHASAN**

4.1 Raw Data	40
4.2 Hasil Ekstraksi Data	41
4.3 Visualisasi Dataset	43
4.4 Hasil Validasi <i>Principal Component Analysis</i> (PCA)	48
4.4.1 Hasil Validasi Reduksi Data 20 Komponen	48
4.4.2 Hasil Validasi Reduksi Data 30 Komponen	49
4.4.3 Hasil Validasi Reduksi Data 40 Komponen	49
4.5 Hasil Validasi <i>Artificial Neural Network</i> (ANN)	50
4.5.1 Hasil Validasi ANN 20 Komponen	51
4.5.2 Hasil Validasi ANN 30 Komponen	57
4.5.3 Hasil Validasi ANN 40 Komponen	64
4.6 Analisa Akurasi dan Loss Pada ANN Dengan Data PCA	71
4.7 Peningkatan Sistem Classification Penelitian Diusulkan dan Penelitian Sebelumnya	73
4.8 Perbandingan Classification Menggunakan PCA dan Tanpa Menggunakan PCA	73
4.8.1 Classification Tanpa Menggunakan PCA	73
4.8.2 Classification Dengan Menggunakan PCA	76

## **BAB V KESIMPULAN DAN SARAN**

5.1 Kesimpulan	79
5.2 Saran	79
Daftar Pustaka	80

## DAFTAR GAMBAR

Gambar 2.1 Arsitektur <i>Auto-Encoder</i> .....	11
Gambar 2.2 Model Arsitektur <i>Artificial Neural Network</i> .....	13
Gambar 2.3 Fungsi Aktivasi <i>Sigmoid</i> .....	15
Gambar 2.4 Fungsi Aktivasi <i>Tanh</i> .....	16
Gambar 2.5 Fungsi Aktivasi <i>ReLU</i> .....	16
Gambar 3.1 <i>Flowchart</i> Alur Langkah-langkah Penelitian .....	22
Gambar 3.2 <i>The Network Architecture</i> UNB .....	24
Gambar 3.3 Alur Ekstraksi Data .....	25
Gambar 3.4 Penggabungan Data <i>malware</i> dan Normal .....	25
Gambar 3.5 Proses Penggabungan Dataset .....	26
Gambar 3.6 <i>Flowchart</i> PCA.....	29
Gambar 3.7 Arsitektur ANN .....	31
Gambar 3.8 <i>Flowchart</i> ANN.....	32
Gambar 4.1 Tampilan PCAP <i>malware</i> pada <i>wireshark</i> .....	40
Gambar 4.2 Tampilan PCAP normal pada <i>wireshark</i> .....	41
Gambar 4.3 Hasil ekstraksi data menjadi csv.....	41
Gambar 4.4 Visualisasi <i>protocol</i> pada data <i>malware</i> .....	43
Gambar 4.5 Visualisasi <i>protocol</i> pada data normal .....	44
Gambar 4.6 Visualisasi <i>protocol</i> pada data gabungan .....	45
Gambar 4.7 Visualisasi jumlah <i>malware</i> dan data normal.....	46
Gambar 4.8 Frekuensi data <i>malware</i> dan normal berdasarkan <i>protocol</i> .....	47
Gambar 4.9 Hasil PCA 20 komponen .....	49
Gambar 4.10 Hasil PCA 30 komponen .....	49
Gambar 4.11 Hasil PCA 40 komponen .....	50
Gambar 4.12 Model akurasi adadelta 20 komponen .....	51
Gambar 4.13 Model loss adadelta 20 komponen .....	51
Gambar 4.14 Confusion Matrix adadelta 20 komponen .....	52
Gambar 4.15 Model akurasi adam 20 komponen.....	53
Gambar 4.16 Model loss adam 20 komponen .....	53
Gambar 4.17 Confusion Matrix adam 20 komponen .....	54
Gambar 4.18 Model akurasi SGD 20 komponen .....	55

Gambar 4.19 Model loss SGD 20 komponen.....	56
Gambar 4.20 Confusion Matrix SGD 20 komponen.....	57
Gambar 4.21 Model akurasi adadelta 30 komponen.....	58
Gambar 4.22 Model loss adadelta 30 komponen.....	58
Gambar 4.23 Confusion Matrix adadelta 30 komponen.....	59
Gambar 4.24 Model akurasi adam 30 komponen.....	60
Gambar 4.25 Model loss adam 30 komponen.....	60
Gambar 4.26 Confusion Matrix adam 30 komponen.....	61
Gambar 4.27 Model akurasi SGD 30 komponen.....	62
Gambar 4.28 Model loss SGD 30 komponen.....	62
Gambar 4.29 Confusion Matrix SGD 30 komponen.....	63
Gambar 4.30 Model akurasi adadelta 40 komponen.....	64
Gambar 4.31 Model loss adadelta 40 komponen.....	65
Gambar 4.32 Confusion Matrix adadelta 40 komponen.....	66
Gambar 4.33 Model akurasi adam 40 komponen.....	67
Gambar 4.34 Model loss adam 40 komponen.....	67
Gambar 4.35 Confusion Matrix adam 40 komponen.....	68
Gambar 4.36 Model akurasi SGD 40 komponen.....	69
Gambar 4.37 Model loss SGD 40 komponen.....	69
Gambar 4.38 Confusion Matrix SGD 40 komponen.....	70
Gambar 4.39 Keseluruhan akurasi dan loss pada classification.....	72
Gambar 4.40 Model akurasi tanpa PCA.....	73
Gambar 4.41 Model loss tanpa PCA.....	74
Gambar 4.42 Confusion Matrix tanpa PCA.....	75
Gambar 4.43 Model akurasi SGD 40 komponen.....	76
Gambar 4.44 Model loss SGD 40 komponen.....	76
Gambar 4.45 Confusion Matrix SGD 40 komponen.....	77

## DAFTAR TABEL

Tabel 1. Penelitian klasifikasi <i>malware</i> 5 tahun terakhir .....	7
Tabel 2. <i>Malware</i> dan jenis keluarganya.....	10
Tabel 3. Tabel <i>Confusion Matrix</i> .....	19
Tabel 4. ANN <i>tuning</i> .....	34
Tabel 5. Model ANN yang digunakan .....	35
Tabel 6. Pembagian Data Pelatihan dan Validasi.....	35
Tabel 7. Komponen parameter reduksi dimensi.....	36
Tabel 8. Struktur Validasi ANN.....	37
Tabel 9. Tabel kebenaran <i>Confusion Matrix</i> .....	38
Tabel 10. Hasil persentasi jumlah <i>protocol</i> pada data <i>malware</i> .....	44
Tabel 11. Hasil persentasi jumlah <i>protocol</i> pada data normal .....	45
Tabel 12. Hasil persentasi jumlah <i>protocol</i> pada data gabungan .....	46
Tabel 13. Hasil jumlah <i>malware</i> dan data normal pada data gabungan.....	47
Tabel 14. Hasil Performa Validasi ANN adadelta 20 Komponen .....	52
Tabel 15. Nilai Confusion Matrix ANN adadelta 20 Komponen.....	53
Tabel 16. Hasil Performa Validasi ANN adam 20 Komponen .....	54
Tabel 17. Nilai Confusion Matrix ANN adam 20 Komponen .....	55
Tabel 18. Hasil Performa Validasi ANN SGD 20 Komponen.....	56
Tabel 19. Nilai Confusion Matrix ANN SGD 20 Komponen.....	57
Tabel 20. Hasil Performa Validasi ANN adadelta 30 Komponen .....	58
Tabel 21. Nilai Confusion Matrix ANN adadelta 20 Komponen.....	59
Tabel 22. Hasil Performa Validasi ANN adam 30 Komponen .....	61
Tabel 23. Nilai Confusion Matrix ANN adam 30 Komponen .....	62
Tabel 24. Hasil Performa Validasi ANN SGD 30 Komponen.....	63
Tabel 25. Nilai Confusion Matrix ANN SGD 30 Komponen.....	64
Tabel 26. Hasil Performa Validasi ANN adadelta 40 Komponen .....	65
Tabel 27. Nilai Confusion Matrix ANN adadelta 40 Komponen.....	66
Tabel 28. Hasil Performa Validasi ANN adam 40 Komponen .....	67
Tabel 29. Nilai Confusion Matrix ANN adam 40 Komponen .....	68
Tabel 30. Hasil Performa Validasi ANN SGD 40 Komponen.....	70
Tabel 31. Nilai Confusion Matrix ANN SGD 40 Komponen.....	71

Tabel 32. Hasil Performa Validasi ANN tanpa PCA .....	74
Tabel 33. Nilai Confusion Matrix ANN tanpa PCA .....	75
Tabel 34. Hasil Performa Validasi ANN SGD 40 Komponen.....	77
Tabel 31. Nilai Confusion Matrix ANN SGD 40 Komponen.....	78



## **DAFTAR LAMPIRAN**

- Lampiran 1. Biodata Diri
- Lampiran 2. USEPT
- Lampiran 3. Grafik Percobaan
- Lampiran 4. Confusion Matriks

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

*Malware* adalah sebuah *software* yang bisa menyusup ke sebuah sistem operasi yang dapat merusak sistem serta mencuri *file-file* pribadi penting yang ada pada system yang disusupi. Beberapa oknum menggunakan *malware* untuk mendapatkan data-data penting dari korban dengan banyak tujuan. *Smartphone* dengan sistem operasi *Android* pun juga menjadi target serangan oknum yang memanfaatkan *malware*. Salah satunya *Adware* yang dapat menyerang *smartphone* dengan melalui iklan yang ada pada aplikasi yang dijalankan ataupun pada *website* yang kita telusuri.

Iklan digunakan untuk mempromosikan, menjual produk atau layanan. Dengan munculnya internet strategi pemasaran bergerak cepat menuju periklanan digital, dan tren ini tidak menunjukkan tanda-tanda mereda. Iklan digital muncul di situs web *online* ataupun aplikasi seluler. Saat ini, kebanyakan *smartphone* menggunakan *platform android*, yang sudah mendorong pertumbuhan ekosistem aplikasi *Android* dengan sangat pesat. Akibatnya, *malware* dapat digunakan untuk melakukan aktivitas berbahaya saat menggunakan hak istimewa yang diberikan oleh sistem operasi *Android*[1]. Penyedia iklan membayar pengembangan aplikasi dan situs web dengan jumlah tertentu untuk setiap tayangan iklan yang ditampilkan kepada pengguna. Perangkat lunak yang melakukan penipuan berupa iklan disebut dengan *adware*[2].

*Adware* adalah suatu iklan yang difungsikan untuk menguntungkan semua pihak, tetapi sekarang *adware* lebih diartikan sebagai penyedia dan sumber utama dari *malware*. *Adware* umumnya dikenal sebagai “perangkat lunak yang secara otomatis menampilkan atau mengunduh materi iklan ketika pengguna sedang online ataupun menjalankan aplikasi[3]. Jenis serangan *adware* berbahaya yang paling sering ialah menyusup serta mengontrol perangkat *Android* pengguna menggunakan *adware* berbahaya[4]. *Adware* juga dapat menyertakan aktivitas berbahaya lainnya seperti, mencuri informasi *sensitive* dari pengguna dan

menyebarkannya kepada pihak ketiga. *Adware* akan terus menerus menampilkan iklan bahkan ketika pengguna tidak menjalankan program yang dikehendaki sebelumnya[5].

Pengembangan *malware* telah menjadi aktivitas yang serius akhir-akhir ini karena jumlah *platform* target meningkat dari tahun ke tahun, yang secara signifikan meningkatkan pengembangan teknik yang memadai untuk mendeteksi *malware*[6]. *Adware* juga menjadi dasar untuk *me-rooting smartphone* pengguna agar mendapatkan akses ke perangkat[2]. Ada dua cara penipuan iklan yaitu pada seluler dan web. Pada seluler iklan akan tampil melalui aplikasi *Android* yang telah berkerjasama dengan iklan tersebut ataupun aplikasi yang telah disisipkan *malware*, sedangkan pada web akan muncul ditampilan atau halaman web dan ada juga saat kita akan mengklik sesuatu tetapi yang muncul adalah iklan.

Ini adalah bukti bahwa *malware* merupakan ancaman yang serius dalam sistem dan sering kali melampaui kapasitas analisis *malware*. Oleh karena itu, diperlukan upaya yang cukup besar untuk melindungi dari serangan *malware*[6]. Penelitian mengenai klasifikasi *malware* sudah banyak dilakukan dalam beberapa tahun ini. Sehingga penelitian mengenai klasifikasi *malware* bukan hal yang baru lagi, namun pengembangan penelitian ini masih terus berjalan hingga sekarang. Hal ini dapat dilihat dari seberapa banyaknya penelitian yang telah dilakukan sebelumnya. Pada penelitian[7] mereka menggunakan algoritma *Random Forest* (RF), *K-Nearest Neighbor* (KNN), *Decision Tree* (DT), *Random Tree* (RT), *Regression* (R) dengan tingkat akurasi sebesar 94%. Dari penelitian tersebut masih memiliki kekurangan dengan proses memakan waktu yang lama dan banyak data yang terlewatkan. Sehingga perlu dilakukan penelitian yang lebih baik dan tidak membutuhkan waktu lama untuk klasifikasi *malware Android* agar dapat mengatasi masalah diatas.

## **1.2. Rumusan Masalah**

Rumusan masalah yang akan dibahas pada tugas akhir ini adalah :

1. Bagaimana mengklasifikasi *malware Android* dengan menggunakan algoritma ANN?

2. Bagaimana menerapkan sistem klasifikasi *Binary* dengan menggunakan algoritma ANN?
3. Bagaimana algoritma PCA mendukung sistem klasifikasi *malware Android* menggunakan ANN?

### **1.3. Batasan Masalah**

Batasan masalah untuk tugas akhir ini sebagai berikut :

1. Dataset yang pakai pada penelitian ini berasal dari *University of New Brunswick* yaitu CIC&Mal2017[8]
2. Algoritma yang digunakan pada klasifikasi *malware Android* adalah algoritma ANN
3. Proses *dimentionality reduction* yang digunakan adalah *Principal Component Analysis*
4. Dalam penelitian ini tidak membahas deteksi dan pencegahan

### **1.4. Tujuan**

Adapun tujuan yang hendak dicapai dari dilakukannya penelitian ini adalah :

1. Menerapkan algoritma ANN untuk klasifikasi data *malware Android*
2. Menerapkan sistem klasifikasi *Binary* dan melakukan peningkatan nilai akurasi dari penelitian sebelumnya
3. Mengetahui pengaruh algoritma *Principal Component Analysis* pada kinerja sistem klasifikasi *malware Android*

### **1.5. Manfaat**

Adapun beberapa manfaat yang didapat dari dilakukannya penelitian ini adalah :

1. Dapat mengklasifikasi *malware Android* dengan digunakan algoritma ANN
2. Algoritma *Principal Component Analysis* sangat berpengaruh untuk melakukan peningkatan kinerja system klasifikasi dengan digunakan algoritma ANN
3. Hasil yang telah didapatkan dari penelitian ini bisa menjadi referensi untuk peningkatan nilai akurasi, spesifitas, sensitivitas, F1-score dan presisi untuk

melakukan klasifikasi dengan menerapkan algoritma *Principal Component Analysis* dan ANN

### **1.6. Metodologi Penelitian**

Penelitian ini akan melewati beberapa tahapan :

1. Tahap Pertama (Studi Pustaka / Literatur)  
Tahap ini dilakukan dengan cara mencari dan membaca literatur dan referensi tentang “Klasifikasi *malware* menggunakan ANN” sehingga dapat menunjang penulisan laporan Tugas Akhir.
2. Tahap Kedua (Prosesing perubahan data PCAP ke csv menggunakan CICFlowMeter )  
Tahap ini dilakukan untuk mempersiapkan data yang akan digunakan.
3. Tahap Ketiga (Pembersihan File)  
Pada tahap ini dilakukan proses dengan melakukan reduksi dimensi
4. Tahap Keempat (Eksperimen)  
Pada tahap ini menyederhanakan data agar lebih maksimum, kemudian melakukan klasifikasi data normal dan malware.
5. Tahap Kelima (Penarikan Kesimpulan dan Saran)  
Pada tahap ini bisa ditarik kesimpulan hasil klasifikasi dan setudi literatur agar untuk penulis selanjutnya yang akan menjadikan bahan referensi.

### **1.7. Sistematika Penulisan**

Agar memudahkan untuk menyusun tugas akhir ini serta memperjelaskan isi dari tiap bab yang ada pada laporan ini, dibuatlah penulisan sistematik dari penelitian ini sebagai berikut.

## **BAB I PENDAHULUAN**

Bab ini berisikan uraian latar belakang dengan sistematik topik yang diambil

## **BAB II TINJAUAN PUSTAKA**

Bab ini berisikan tentang literatur yang relevan, bagan teori dan bagan berfikir

### **BAB III METODOLOGI**

Bab ini menjelaskan step by step dan terperinci tentang apa yang akan dilakukan dan digunakan dengan mencari kemudian menganalisa tema yang sudah dibuat penulis untuk tugas akhir.

### **BAB IV PENGUJIAN DAN ANALISA**

Bab ini menjelaskan dari hasil pengujian yang sudah dilakukan dan di analisa dengan hasil klasifikasi yang sudah dibuat.

### **BAB V KESIMPULAN**

Bab ini berisikan tentang kesimpulan yang telah didapatkan oleh penulis serta merupakan jawaban dari semua tujuan yang akan dicapai pada bab I (Pendahuluan).

## DAFTAR PUSTAKA

- [1] A. Pektaş and T. Acarman, “Ensemble machine learning approach for android malware classification using hybrid features,” in *International Conference on Computer Recognition Systems*, 2017, pp. 191–200.
- [2] S. Suresh, F. Di Troia, K. Potika, and M. Stamp, “An analysis of Android adware,” *J. Comput. Virol. Hacking Tech.*, vol. 15, no. 3, pp. 147–160, 2019.
- [3] J. Gao, L. Li, P. Kong, T. F. Bissyandé, and J. Klein, “Should you consider adware as malware in your study?,” in *2019 IEEE 26th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, 2019, pp. 604–608.
- [4] K. Lee and H. Park, “Malicious Adware Detection on Android Platform using Dynamic Random Forest,” in *International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, 2019, pp. 609–617.
- [5] J. Y. Ndagi and J. K. Alhassan, “Machine Learning Classification Algorithms for Adware in Android Devices: A Comparative Evaluation and Analysis,” in *2019 15th International Conference on Electronics, Computer and Computation (ICECCO)*, 2019, pp. 1–6.
- [6] F. O. Catak, A. F. Yazıcı, O. Elezaj, and J. Ahmed, “Deep learning based Sequential model for malware analysis using Windows exe API Calls,” *PeerJ Comput. Sci.*, vol. 6, p. e285, 2020.
- [7] M. Murtaz, H. Azwar, S. B. Ali, and S. Rehman, “A framework for Android Malware detection and classification,” in *2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, 2018, pp. 1–5.
- [8] 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,” in *2018 International Carnahan Conference on Security Technology (ICCST)*, 2018, pp. 1–7.
- [9] K. H. Walse, R. V Dharaskar, and V. M. Thakare, “PCA Based Optimal ANN Classifiers for Human Activity Recognition Using Mobile Sensors Data,” in *Proceedings of First International Conference on Information and Communication Technology*, vol. 1, pp. 429-436, 2016.
- [10] M. K. Alzaylaee, S. Y. Yerima, and S. Sezer, “DL-Droid: Deep learning based android malware detection using real devices,” *Comput. Secur.*, vol. 89, p. 101663, 2020.
- [11] O. T. Suryati and A. Budiono, “Impact Analysis of Malware Based on Call Network API with Heuristic Detection Method,” *Int. J. Adv. Data Inf. Syst.*, vol. 1, no. 1, pp. 1–8, 2020.
- [12] M. B. B. de Robles, J. A. C. Hermocilla, and J. P. Pabico, “Characterization

- and Classification of Malware Traffic over the Tor Network.” in *Proceedings of the 20th Philippine Computing Science Congress (PCSC 2020)*, pp. 78-87, 2020.
- [13] J. Eliyanto, “Reduksi Dimensi untuk Meningkatkan Performa Metode Fuzzy Klastering pada Big Data,” in *Public Knowledge Project*, vol. 1, no. 1, pp. 27–36, 2019.
- [14] H. Lu, S. Liu, H. Wei, and J. Tu, “Expert Systems with Applications Multi-kernel fuzzy clustering based on auto-encoder for fMRI functional network,” *Expert Systems with Applications*, vol. 159, 2020.
- [15] P. Hartono, “Mixing Autoencoder With Classifier: Conceptual Data Visualization,” *IEEE Access*, vol. 8, no. 1, pp. 105301-105310, 2020.
- [16] D. Arivudainambi, V. K. KA, P. Visu, and others, “Malware traffic classification using principal component analysis and artificial neural network for extreme surveillance,” *Comput. Commun.*, vol. 147, pp. 50–57, 2019.
- [17] Y.-M. Kwon, J.-J. An, M.-J. Lim, S. Cho, and W.-M. Gal, “Malware Classification Using Simhash Encoding and PCA (MCSP),” *Symmetry (Basel)*, vol. 12, no. 5, p. 830, 2020.
- [18] R. S. Suhartanto, C. Dewi, and L. Muflikhah, “Implementasi Jaringan Syaraf Tiruan Backpropagation untuk Mendiagnosis Penyakit Kulit pada Anak,” *Jurnal Masyarakat Informatika (JMASIF)*, vol. 1, no. 7, pp. 555–562, 2017.
- [19] I. M. Nasser, M. Al-Shawwa, and S. S. Abu-Naser, “Artificial Neural Network for Diagnose Autism Spectrum Disorder,” in *International Journal of Academic Information System Research (IJASIR)*, vol. 3, 2019.
- [20] S. A. Aditya, “Inovasi Metode ANN-S Untuk Pengecekan Baterai Berbasis Labview,” in *Seminar Nasional Teknik Elektro*, 2020, vol. 5, no. 2, pp. 235–238.
- [21] J. Teknovasi, E. B. Nababan, and M. Zarlis, “BIPOLAR DALAM ALGORITMA BACKPROPAGATION PADA PREDIKSI KEMAMPUAN SISWA,” *Jurnal Ilmiah Informatika (JIF)*, vol. 02, pp. 103–116, 2015.
- [22] Y. Yu *et al.*, “RMAF: Relu-Memristor-Like Activation Function for Deep Learning,” *IEEE Access*, vol. 8, pp. 72727–72741, 2020.
- [23] T. K. R. Arvind, M. Brand, C. Heidorn, S. Boppu, F. Hannig, and J. Teich, “Hardware Implementation of Hyperbolic Tangent Activation Function for Floating Point Formats,” *IEEE Access*, p. 121, 2020.
- [24] R. Adam, D. O. Melinte, and L. Vladareanu, “Facial Expressions Recognition for Human – Robot Interaction Using Deep Convolutional Neural,” *Sensors*, vol. 20, 2020.
- [25] V. N. Sewdien, R. Preece, J. L. R. Torres, E. Rakhshani, and M. Van Der Meijden, “Assessment of critical parameters for arti fi cial neural networks based short-term wind generation forecasting,” *Renew. Energy*, vol. 161, pp.



878–892, 2020.

- [26] K. Rahman, “Training Sensitivity in Graph Isomorphism Network,” *ACM Digital Library* pp. 2181–2184, 2020.
- [27] G. Perin and S. Picek, “On the Influence of Optimizers in Deep Learning-based Side-channel Analysis,” *Cryptology ePrint Archive*, pp. 13–28, 2020.
- [28] I. Kandel and M. Castelli, “Comparative Study of First Order Optimizers for Image Classification Using Convolutional Neural Networks on Histopathology Images,” *J. Imagin*, vol. 6, 2020.
- [29] Q. Wang, Y. Ma, K. Zhao, and Y. Tian, “A Comprehensive Survey of Loss Functions in Machine Learning,” *Ann. Data Sci*, 2020.
- [30] H. Qin, R. Gong, X. Liu, X. Bai, J. Song, and N. Sebe, “Binary neural networks : A survey,” *Pattern Recognit*, vol. 105, 2020.'
- [31] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, “The impact of class imbalance in classification performance metrics based on the binary confusion matrix,” *Pattern Recognit.*, vol. 91, pp. 216–231, 2019.
- [32] A. H. Lashkari, A. F. A. Kadir, H. Gonzalez, K. F. Mbah, and A. A. Ghorbani, “Towards a network-based framework for android malware detection and characterization,” in *2017 15th Annual Conference on Privacy, Security and Trust (PST)*, 2017, pp. 233–23309.
- [33] M. Noorani, S. Mancoridis, and S. Weber, “*On the Detection of Malware on Virtual Assistants Based on Behavioral Anomalies*,” *Drexel University*, 2019.