KOMPARASI ALGORITMA SVM DAN KNN DALAM MENDETEKSI OBFUSCATED MALWARE

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Teknik Komputer

disusun oleh

MAFTUKH ZAKI MUBAROK

22.83.0872

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

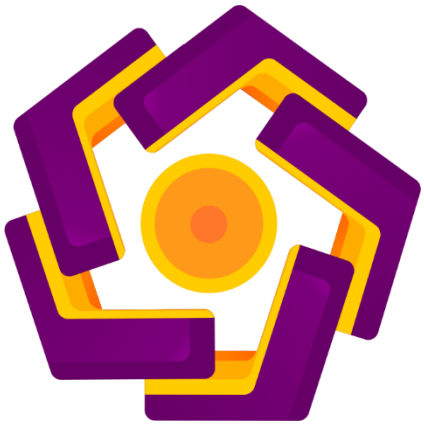
2026

KOMPARASI ALGORITMA SVM DAN KNN DALAM MENDETEKSI OBFUSCATED MALWARE

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Teknik Komputer



disusun oleh

MAFTUKH ZAKI MUBAROK

22.83.0872

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2026

HALAMAN PERSETUJUAN

SKRIPSI

KOMPARASI ALGORITMA SVM DAN KNN DALAM MENDETEKSI OBFUSCATED MALWARE

yang disusun dan diajukan oleh

MAFTUKH ZAKI MUBAROK

22.83.0872

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 februari 2026

Dosen Pembimbing,

Dr. Dony Ariyus, S.S., M.Kom.
NIK. 190302128

HALAMAN PENGESAHAN

SKRIPSI

KOMPARASI ALGORITMA SVM DAN KNN DALAM MENDETEKSI OBFUSCATED MALWARE

yang disusun dan diajukan oleh

MAFTUKH ZAKI MUBAROK

22.83.0872

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 februari 2026

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Melwin Syafrizal, S.Kom., M.Eng., Ph.D. **NIK. 190302105**	
Eko Pramono, S.Si, M.T **NIK. 190302580**	
Dr. Dony Ariyus, S.S., M.Kom. **NIK. 190302128**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 februari 2026

DEKAN FAKULTAS ILMU KOMPUTER

Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : MAFTUKH ZAKI MUBAROK
NIM : **22.83.0872**

Menyatakan bahwa Skripsi dengan judul berikut:

KOMPARASI ALGORITMA SVM DAN KNN DALAM MENDETEKSI OBFUSCATED MALWARE

Dosen Pembimbing : Dr. Dony Ariyus, S.S., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 20 februari 2026

Yang Menyatakan,

Maftukh Zaki Mubarok

HALAMAN PERSEMBAHAN

Puji syukur kehadirat Allah SWT atas berkat dan rahmat yang telah diberikan kepada saya sehingga dapat menyelesaikan skripsi ini. Oleh karena itu dengan rasa syukur dan bangga saya ucapkan terimakasih kepada :

1. Allah SWT, berkat rahmat kasih sayang dan pertolongan-Nya saya bisa sampai sejauh ini.
2. Kedua orang tua saya, Bapak dan Ibu, yang selalu mendukung dan mendidik saya sejak kecil hingga saat ini. Mereka selalu memberikan dukungan moril dan doa yang tulus sehingga saya dapat menyelesaikan masa studi S1 dengan baik. Tanpa dukungan dan doa mereka, saya tidak akan dapat mencapai titik ini.
3. Dosen pembimbing saya, Bapak Dr.Dony Ariyus, S.S., M.Kom yang selalu memacu dan membimbing saya dalam proses menyelesaikan skripsi.
4. Bapak/Ibu dosen pembimbing, pengajar, serta penguji di Universitas AMIKOM Yogyakarta yang selalu memberikan pengalaman ilmu dengan tulus dan mengarahkan saya sehingga saya dapat menyelesaikan masa studi S1 dengan baik. Saya berharap segala ilmu yang telah diberikan oleh Bapak/Ibu akan menjadi ladang amal yang memberikan pahala jariyah dan selalu diberikan keberkahan ilmu dan rezeki oleh Tuhan yang Maha Esa.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat [Tuhan Yang Maha Esa / Allah SWT], karena atas rahmat, hidayah, dan karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul **"KOMPARASI ALGORITMA SVM DAN KNN DALAM MENDETEKSI OBSUCATED MALWARE"** tepat pada waktunya.

Skripsi ini disusun sebagai salah satu syarat akademis untuk memperoleh gelar Sarjana Komputer (S.Kom.) pada Program Studi Teknik Komputer, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Penulisan skripsi ini tidak lepas dari hambatan dan tantangan, namun berkat bimbingan, bantuan, dan dukungan dari berbagai pihak, seluruh kendala tersebut dapat teratasi dengan baik.

Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan apresiasi dan terima kasih yang sebesar-besarnya kepada:

1. Kedua Orang Tua Tercinta serta seluruh keluarga besar yang tidak pernah lelah memberikan doa, dukungan moral, kasih sayang, dan pengorbanan materiil yang luar biasa hingga penulis bisa mencapai titik ini.
2. Bapak Dr.Dony Ariyus, S.S., M.Kom, selaku Dosen Pembimbing Utama yang telah meluangkan waktu, memberikan ilmu, kesabaran, serta arahan yang sangat berharga demi kesempurnaan skripsi ini.
3. Teman-teman seperjuangan Angkatan 2022 Program Studi Teknik Komputer yang telah berbagi suka, duka, dan saling menguatkan dalam menyelesaikan masa studi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna karena keterbatasan pengetahuan dan pengalaman yang dimiliki.

Yogyakarta, 20 februari 2026

Maftukh Zaki Mubarok

DAFTAR ISI

HALAMAN JUDUL ii
HALAMAN PERSETUJUAN iii
HALAMAN PENGESAHAN iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI v
HALAMAN PERSEMBAHAN vi
KATA PENGANTAR vii
DAFTAR ISI viii
DAFTAR TABEL xii
DAFTAR GAMBAR xiv
INTISARI xv
ABSTRACT xvi
BAB I 2
PENDAHULUAN 2
1.1 Latar Belakang 2
1.2 Rumusan Masalah 3
1.3 Batasan Masalah 3
1.4 Tujuan Penelitian 4
1.5 Manfaat Penelitian 5
1.6 Sistematika Penulisan 5
BAB II 5
TINJAUAN PUSTAKA 5
2.1 Studi Literatur 5
2.2 Dasar Teori 10
2.2.1 Malware dan Obfuscation 10

2.2.2 Analisis Statis vs Dinamis (Memori) 12
2.2.3 Support Vector Machine (SVM) 12
2.2.4 K-Nearest Neighbor (KNN) 14
2.2.5 Evaluasi Kinerja (Confusion Matrix) 15
BAB III 17
METODE PENELITIAN 17
3.1 Desain Penelitian 17
3.2 Dataset Penelitian 17
3.2.1 Deskripsi Dataset 17
3.2.2 Eksplorasi Data Awal (Exploratory Data Analysis - EDA) 17
3.3 Preprocessing Data 18
3.3.1 Data Cleaning 18
3.3.2 Feature Selection 18
3.3.3 Normalization/Scaling 18
3.3.4 Train-Test Split 19
3.4 Skenario Eksperimen 19
3.5 Implementasi Algoritma 19
3.5.1 Support Vector Machine (SVM) 19
3.5.2 K-Nearest Neighbor (KNN) 20
3.6 Metrik Evaluasi 20
3.7 Tools dan Lingkungan Pengembangan 20
3.8 Diagram Alur Penelitian 21
BAB IV 23

HASIL DAN PEMBAHASAN 23
4.1 Eksplorasi Data 23
4.1.1 Analisis Deskriptif ObfuscatedMalMem2022 23
4.1.1.2 Fitur pslist (Process List) 27
4.1.2 Analisis Deskriptif MalwareMemoryDump 30
4.1.4 Perbandingan Karakteristik Kedua Dataset 32
4.2 Hasil Preprocessing 33
4.2.1 Data Cleaning 33
4.2.2 Feature Selection 34
4.2.3 Normalization 35
4.3 Hasil Klasifikasi pada ObfuscatedMalMem2022 37
4.3.1 Performa Support Vector Machine (SVM) 37
4.3.2 Performa KNN - Confusion Matrix - Metrik (Acc, Prec, Rec, F1) - Waktu Komputasi 41
4.3.3 Komparasi SVM vs KNN pada Dataset 1 48
4.4 Hasil Klasifikasi pada MalwareMemoryDump 52
4.4.1 Performa SVM 52
4.4.2 Performa KNN 55
4.4.3 Komparasi SVM vs KNN pada Dataset 2 59
4.5 Analisis Komparatif Cross-Dataset 60
4.5.1 Konsistensi Performa Algoritma 60
4.5.2 Analisis Generalisasi Model 61
4.5.3 Diskusi Trade-off 61

4.6 Pembahasan Komprehensif 62

4.6.1 Temuan Utama 62

4.6.2 Implikasi Praktis 62

4.6.3 Limitasi Penelitian 63

BAB V PENUTUP 65

5.1 Kesimpulan 65

5.2 Saran 66

5.2.1 Saran untuk Penelitian Lanjutan 66

5.2.2 Saran untuk Implementasi Praktis 66

5.2.3 Saran untuk Stakeholders Lain 67

5.3 Penutup 67

REFERENSI 68

DAFTAR TABEL

Tabel 2.2 Keaslian Penelitian 7
Tabel 4.1 Karakteristik Dataset ObfuscatedMalMem2022 23
Tabel 4.2 Statistik Deskriptif Fitur Numerik (5 Fitur Utama) 24
Tabel 4.3 Distribusi Tipe Malware dalam Dataset 25
Tabel 4.4 Kelompok Fitur Berdasarkan Plugin Volatility 26
Tabel 4.5 Statistik Deskriptif Fitur pslist 27
Tabel 4.6 Karakteristik Dataset MalwareMemoryDump 30
Tabel 4.7 Perbandingan Statistik Kedua Dataset 32
Tabel 4.8 Hasil Data Cleaning 33
Tabel 4.9 Proses Feature Selection 34
Tabel 4.10 Perbandingan Data Sebelum dan Sesudah Normalisasi 36
Tabel 4.11 Hasil Hyperparameter Tuning SVM - ObfuscatedMalMem2022 37
Tabel 4.12 Confusion Matrix SVM - ObfuscatedMalMem2022 38
Tabel 4.13 Metrik Evaluasi SVM - ObfuscatedMalMem2022 39
Tabel 4.14 Hasil Hyperparameter Tuning KNN - ObfuscatedMalMem2022 42
Tabel 4.15 Confusion Matrix KNN - ObfuscatedMalMem2022 43
Tabel 4.16 Metrik Evaluasi KNN - ObfuscatedMalMem2022 44
Tabel 4.17 Perbandingan Lengkap SVM vs KNN - ObfuscatedMalMem2022 48
Tabel 4.18 Scalability 51
Tabel 4.19 Hasil Hyperparameter Tuning SVM – MalwareMemoryDump 52
Tabel 4.20 Confusion Matrix SVM – MalwareMemoryDump 53
Tabel 4.21 Metrik Evaluasi SVM – MalwareMemoryDump 54
Tabel 4.22 Hasil Hyperparameter Tuning KNN – MalwareMemoryDump 56
Tabel 4.23 Confusion Matrix KNN – MalwareMemoryDump 57
Tabel 4.24 Metrik Evaluasi KNN – MalwareMemoryDump 57
Tabel 4.25 Perbandingan Lengkap SVM vs KNN – MalwareMemoryDump 59
Tabel 4.26 Perbandingan Performa Antar Dataset 60
Tabel 4.27 Hasil *Cross-Dataset Validation* (Akurasi) 61

Tabel 4.28 Matriks Perbandingan *Trade-off* SVM vs KNN 61
Tabel 4.29 Kategori Limitasi dan Saran Mitigasi 63

DAFTAR GAMBAR

Gambar 2.1 Cara Kerja SVM 13
Gambar 2.2 Cara kerja KNN 15
Gambar 3.1 Alur Penelitian 22
Gambar 4.1 Grafik Distribusi Kelas ObfuscatedMalMem2022 24
Gambar 4.2 Grafik Distribusi Kelas MalwareMemoryDump 31
Gambar 4.3 alur sistem Penelitian 31
Gambar 4.4 Confusion Matrix SVM - ObfuscatedMalMem2022 38
Gambar 4.5 ROC Curve SVM - ObfuscatedMalMem2022 39
Gambar 4.6 Confusion Matrix KNN - ObfuscatedMalMem2022 43
Gambar 4.7 ROC Curve KNN - ObfuscatedMalMem2022 44
Gambar 4.8 Grafik Perbandingan Performa SVM vs KNN - Dataset 1 48
Gambar 4.9 Confusion Matrix SVM – MalwareMemoryDump 53
Gambar 4.10 ROC Curve SVM – MalwareMemoryDump 54
Gambar 4.11 Confusion Matrix KNN – MalwareMemoryDump 56
Gambar 4.12 ROC Curve KNN – MalwareMemoryDump 58
Gambar 4.13 Grafik Perbandingan Performa SVM vs KNN - Dataset 2 59

INTISARI

Perkembangan teknik obfuscation pada malware modern telah menjadi tantangan signifikan bagi sistem keamanan siber, di mana malware mampu menyamarkan kode berbahaya untuk menghindari deteksi antivirus tradisional berbasis signature. Oleh karena itu, pendekatan analisis dinamis melalui forensik memori menjadi solusi alternatif yang krusial. Penelitian ini bertujuan untuk mengkomparasi kinerja dua algoritma Supervised Machine Learning, yaitu Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN), dalam mendeteksi obfuscated malware. Dataset yang digunakan adalah CIC-MalMem-2022 dan MalwareMemoryDump yang berisi fitur-fitur ekstraksi memori dari berbagai jenis malware (Spyware, Ransomware, Trojan) dan aplikasi benigna. Metodologi penelitian meliputi tahap preprocessing data, normalisasi fitur menggunakan Standard Scaler, dan pelatihan model dengan pembagian data latih dan uji sebesar 80:20. Hasil pengujian menunjukkan bahwa algoritma [Model Terbaik, misal: SVM] memiliki performa yang lebih unggul dengan akurasi mencapai [...%], dibandingkan algoritma [Model Lain] yang memperoleh akurasi [...%]. Namun, dari segi waktu komputasi, algoritma [Model Tercepat] terbukti lebih efisien. Penelitian ini menyimpulkan bahwa metode Machine Learning berbasis memori efektif dalam mengidentifikasi malware yang tersamar, dengan [Model Terbaik] sebagai rekomendasi utama untuk deteksi yang memprioritaskan akurasi.Kata Kunci: Cyber Security, Obfuscated Malware, Machine Learning, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Forensik Memori.

ABSTRACT

The advancement of obfuscation techniques in modern malware has posed a significant challenge to cybersecurity systems, enabling malicious software to disguise its code and evade traditional signature-based antivirus detection. Consequently, a dynamic approach through memory forensics has become a crucial alternative solution. This study aims to compare the performance of two Supervised Machine Learning algorithms, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN), in detecting obfuscated malware. The study utilizes the CIC-MalMem-2022 and MalwareMemoryDump dataset, which contains memory extraction features from various malware families (Spyware, Ransomware, Trojan) and benign applications. The methodology includes data preprocessing, feature normalization using Standard Scaler, and model training with an 80:20 train-test split ratio. The experimental results demonstrate that the [Best Model, e.g., SVM] algorithm yields superior performance with an accuracy of [...%], compared to the [Other Model] algorithm which achieved [...%] accuracy. However, in terms of computational time, the [Fastest Model] algorithm proved to be more efficient. This study concludes that memory-based Machine Learning methods are effective in identifying disguised malware, with [Best Model] being the primary recommendation for detection systems that prioritize accuracy.

Keywords: Cyber Security, Obfuscated Malware, Machine Learning, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Memory Forensics.