BAB I
PENDAHULUAN

1.1 Latar Belakang

Pesatnya perkembangan teknologi informasi dan internet telah membawa dampak positif yang signifikan, namun juga membuka celah keamanan yang serius. Salah satu ancaman terbesar dalam keamanan siber saat ini adalah serangan malware (*malicious software*). Berdasarkan laporan AV-Test Institute (2024), terdapat lebih dari 450.000 varian malware baru yang terdeteksi setiap harinya, dengan tingkat kecanggihan yang semakin tinggi dan metode penyebaran yang semakin kompleks[1].

Metode deteksi malware tradisional yang berbasis tanda tangan (*signature-based detection*) kini dianggap kurang efektif [2]. Hal ini disebabkan oleh kemunculan teknik obfuscation (pengaburan kode), di mana pembuat malware sengaja memodifikasi struktur kode program untuk menyembunyikan logika jahatnya tanpa mengubah fungsi aslinya. Malware jenis ini, yang dikenal sebagai *Obfuscated Malware*, mampu menghindari deteksi antivirus konvensional karena signature atau pola kodenya selalu berubah-ubah meskipun fungsi dasarnya tetap sama.

Untuk mengatasi keterbatasan tersebut, diperlukan pendekatan analisis dinamis, salah satunya melalui analisis forensik memori (*memory analysis*). Ketika *obfuscated malware* dieksekusi, ia harus melakukan "*unpacking*" atau membuka enkripsinya di dalam memori (RAM) agar dapat berjalan. Oleh karena itu, artefak yang tertinggal di memori menjadi sumber data yang lebih akurat dibandingkan analisis file statis, karena kode malware yang sesungguhnya dapat diamati pada tahap runtime.

Penerapan *Machine Learning* (pembelajaran mesin) pada data forensik memori menawarkan solusi menjanjikan untuk mendeteksi pola anomali yang sulit dilihat manusia [3]. Fitur-fitur yang diekstraksi dari analisis memori, seperti daftar proses, modul DLL yang dimuat, dan koneksi jaringan aktif, menghasilkan ruang fitur berdimensi tinggi yang membutuhkan algoritma klasifikasi yang mampu.

menangani kompleksitas tersebut secara efektif. Dua algoritma yang dianggap relevan dan saling kontras karakteristiknya untuk kasus ini adalah *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN). SVM dianggap cocok untuk deteksi *obfuscated malware* karena kemampuannya mencari *hyperplane* optimal pada data berdimensi tinggi hasil ekstraksi memori, sehingga mampu memisahkan pola perilaku malware dari aplikasi normal meskipun kode sumbernya telah disamarkan [4]. Di sisi lain, KNN berpotensi efektif karena malware yang berasal dari keluarga (*family*) yang sama cenderung membentuk kluster berdekatan dalam ruang fitur memori, sehingga pendekatan berbasis kemiripan dengan tetangga terdekat dapat mengidentifikasi ancaman baru yang belum pernah dilihat sebelumnya [5]. Namun, kedua algoritma ini memiliki karakteristik yang berbeda secara fundamental: SVM merupakan pendekatan berbasis margin yang cenderung lebih robust terhadap noise namun membutuhkan waktu komputasi tinggi, sedangkan KNN merupakan pendekatan berbasis instance yang lebih ringan namun sensitif terhadap distribusi data. Perbedaan mendasar ini menimbulkan pertanyaan algoritmik yang belum terjawab secara spesifik: algoritma manakah yang memberikan keseimbangan terbaik antara akurasi deteksi dan efisiensi komputasi pada kasus *obfuscated malware* berbasis analisis forensik memori? Oleh karena itu, komparasi terhadap kedua algoritma ini menjadi penting dilakukan.

Penelitian ini akan berfokus pada komparasi kinerja algoritma SVM dan KNN dalam mendeteksi *obfuscated malware* menggunakan dataset CIC-MalMem-2022 dan *MalwareMemoryDump* [6]. Perbandingan ini penting dilakukan untuk mengetahui algoritma mana yang memberikan keseimbangan terbaik antara akurasi deteksi dan efisiensi waktu komputasi, sehingga dapat menjadi rujukan dalam pengembangan sistem keamanan siber yang lebih cerdas dan responsif terhadap ancaman malware modern.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1 Bagaimana penerapan algoritma *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN) untuk klasifikasi malware pada dataset CIC-MalMem-2022 dan *MalwareMemoryDump*?

2 Bagaimana perbandingan performa algoritma SVM dan KNN dilihat dari parameter *accuracy*, *precision*, *recall*, dan *f1-score*?

3 Algoritma manakah yang lebih efisien dari segi waktu komputasi (*training time*) dalam mendeteksi *obfuscated malware*?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan fokus, penulis menetapkan batasan masalah sebagai berikut:

1. **Dataset**: Data yang digunakan adalah dua dataset berbasis analisis memori yang diperoleh dari Canadian Institute for Cybersecurity (CIC):
   - **ObfuscatedMalMem2022** (Obfuscated Malware Memory Dataset 2022)
   - **MalwareMemoryDump** (Malware Memory Dump Dataset)

Kedua dataset memiliki 58.596 sampel dengan 55+ fitur ekstraksi memori yang saling melengkapi untuk analisis komprehensif.

2. **Algoritma**: Penelitian hanya membandingkan dua algoritma Supervised Learning, yaitu SVM (Support Vector Machine) dan KNN (K-Nearest Neighbor).
3. **Klasifikasi**: Fokus klasifikasi adalah Binary Classification, yaitu mengelompokkan data menjadi dua kelas: Benign (Aman) dan Malware (Berbahaya), tidak sampai mengklasifikasikan jenis famili malware secara spesifik.
4. **Tools**: Implementasi dilakukan menggunakan bahasa pemrograman Python dengan library Scikit-Learn untuk machine learning dan library pendukung lainnya seperti Pandas, NumPy, dan Matplotlib untuk data processing dan visualisasi.

5. **Metrik Evaluasi**: Evaluasi performa menggunakan confusion matrix dengan metrik accuracy, precision, recall, f1-score, serta pengukuran waktu komputasi (training time dan testing time).
6. **Skenario Eksperimen**: Penelitian akan melakukan eksperimen pada kedua dataset secara terpisah untuk membandingkan konsistensi performa algoritma, serta mengeksplorasi kemungkinan cross-dataset validation.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah:

1. Mengimplementasikan algoritma SVM dan KNN sebagai pendekatan alternatif terhadap keterbatasan metode deteksi berbasis tanda tangan (signature-based detection) yang tidak efektif menghadapi obfuscated malware, dengan memanfaatkan fitur-fitur analisis forensik memori dari dataset ObfuscatedMalMem2022 dan MalwareMemoryDump sebagai sumber data yang lebih akurat karena merepresentasikan perilaku malware pada tahap runtime.
2. Mengukur dan menganalisis efektivitas kedua algoritma dalam mendeteksi pola anomali obfuscated malware yang sulit diidentifikasi secara manual, menggunakan metrik accuracy, precision, recall, dan f1-score pada masing-masing dataset untuk memperoleh gambaran komprehensif mengenai kualitas deteksi pada data forensik memori berdimensi tinggi.
3. Menentukan algoritma yang memberikan keseimbangan terbaik antara akurasi deteksi dan efisiensi waktu komputasi, sebagai jawaban atas perbedaan mendasar antara SVM yang berbasis margin dan KNN yang berbasis instance, sehingga dapat menjadi rujukan konkret dalam pengembangan sistem keamanan siber yang cerdas dan responsif terhadap ancaman obfuscated malware modern.
4. Menganalisis konsistensi performa kedua algoritma ketika diuji pada dua dataset forensik memori yang berbeda, guna memastikan bahwa rekomendasi algoritma yang dihasilkan tidak hanya berlaku pada satu kondisi data tertentu melainkan dapat digeneralisasikan sebagai solusi yang

andal dalam menghadapi variasi ancaman obfuscated malware yang terus berkembang.

1.5 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Manfaat Teoretis: Menambah literatur dan wawasan mengenai penerapan *Machine Learning* dalam bidang keamanan siber (*Cyber Security*), khususnya dalam deteksi malware berbasis memori.
2. Manfaat Praktis: Memberikan rekomendasi teknis mengenai pemilihan algoritma yang tepat bagi praktisi keamanan siber dalam membangun sistem deteksi intrusi yang efisien.

1.6 Sistematika Penulisan

BAB 1: PENDAHULUAN Berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB 2: LANDASAN TEORI Membahas teori-teori pendukung seperti Konsep Malware dan Obfuscation, Analisis Memori, *Machine Learning*, algoritma SVM, algoritma KNN, serta tinjauan pustaka (penelitian terdahulu) yang relevan.

BAB 3: METODOLOGI PENELITIAN Menjelaskan tahapan penelitian mulai dari pengumpulan dataset, *preprocessing* data, perancangan model, skenario pengujian, hingga metode evaluasi yang digunakan.

BAB 4: HASIL DAN PEMBAHASAN Menyajikan hasil implementasi program, analisis data, grafik perbandingan performa (*confusion matrix*), serta pembahasan mengenai hasil komparasi kedua algoritma.

BAB 5: PENUTUP Berisi kesimpulan yang diperoleh dari hasil penelitian serta saran-saran untuk pengembangan penelitian selanjutnya.