# BAB I
# PENDAHULUAN

## 1.1. Latar Belakang

Perkembangan sistem komputer yang sangat pesat seringkali membuat banyak orang kewalahan dalam mengikuti perubahan yang terjadi. Setiap hari, sistem digital terus diperbarui, meningkatkan kompleksitas dan kecepatan teknologi yang kita gunakan. Namun, kemajuan ini sering kali bergerak lebih cepat daripada kemampuan manusia untuk mempelajarinya secara menyeluruh, sehingga banyak yang kesulitan memahami dan menggunakan teknologi dengan aman (Von Neumann, 2012). Oleh karena itu, banyak sistem keamanan siber yang dikembangkan dan diimplementasikan saat ini tidak dapat berjalan dengan optimal. Sistem keamanan komputer tidak hanya bekerja sendiri, melainkan juga memerlukan kontribusi dari pemahaman para pengguna untuk menjaga komputer mereka, mulai dari memahami berbagai serangan digital hingga memiliki sikap yang selalu siaga dalam menggunakan berbagai layanan internet.

Salah satu serangan digital yang diuntungkan oleh kurangnya pemahaman banyak pengguna adalah serangan *phishing*, yaitu serangan yang berusaha menipu pengguna dengan mengatasnamakan individu atau organisasi terpercaya (Jansson et al., 2013). Serangan *phishing* bisa menyerang berbagai sistem operasi, bahkan Android (Aonzo et al., 2018). Banyak pengguna tidak memahami masalah keamanan dari serangan *phishing*, sehingga mereka sering termakan godaan atau mempercayai apa yang dikirim oleh penyerang. *Phishing* dapat dilakukan melalui berbagai cara, seperti situs web (Whittaker et al., 2010), UI yang sengaja dibuat membingungkan (Huang et al., 2012), dan dokumen atau link yang dikirim melalui surel (Shankar et al., 2019). Salah satu metode yang semakin populer dalam serangan *phishing* adalah penggunaan dokumen seperti DOCX, PPTX, XLSX, dan PDF. Penyerang menyamarkan *malware* dalam bentuk makro atau skrip berbahaya yang dieksekusi saat pengguna membuka dokumen tersebut (SonicWall, 2022). Dokumen ini sering dikirim sebagai lampiran email dari sumber yang tampak terpercaya,

membuat pengguna lengah dan lebih mungkin untuk membukanya. Keberhasilan serangan ini seringkali karena pengguna tidak waspada terhadap potensi risiko terkait file dokumen yang tampak biasa dan umum digunakan dalam lingkungan kerja sehari-hari. Hasil serangan bisa beragam, mulai dari pencurian informasi penting seperti kartu kredit (Arachchilage et al., 2016) dan identitas pribadi (Eng et al., 2015), hingga penyisipan *malware* ke komputer pengguna (Chaudhry et al., 2016).

Serangan digital terus berkembang dan beradaptasi dengan cepat dalam dunia digital yang selalu berubah (Lin et al., 2019). Hal ini menyulitkan sistem keamanan saat ini karena meskipun terus diperbarui, banyak sistem keamanan masih bergantung pada pola serangan yang sudah dikenali sebelumnya. Proses pembelajaran sistem keamanan digital sering didasarkan pada serangan yang telah terjadi, dan algoritma keamanan hanya dapat mendeteksi ancaman yang serupa dengan yang ada dalam database mereka (Díaz, 2020).

Dari penjelasan tersebut, dapat dilihat bahwa ada beberapa permasalahan yang membuat sistem keamanan digital kurang efektif. Walaupun teknologi terus berubah dan menjadi lebih efektif, masalah keamanan siber tidak selalu diuntungkan dari perkembangan teknologi, karena serangan juga diuntungkan dari perkembangan teknologi yang terus terjadi. Oleh karena itu, penelitian ini akan menganalisis *malware* secara dinamis dengan bantuan algoritma machine learning terhadap *malware* yang didapat dari serangan *phishing* untuk meningkatkan efektivitas keamanan siber. Penelitian ini akan menggunakan laboratorium komputer dalam sebuah komputer virtual dengan sistem operasi Linux untuk menciptakan lingkungan terkendali agar penelitian dilakukan dengan aman, serta melakukan analisis dengan bantuan machine learning melalui metode *decision tree*. Diharapkan hasil analisis ini akan meningkatkan pemahaman tentang proses serangan *malware* dalam sistem komputer.

Machine learning dalam analisis *malware* dapat digunakan dalam berbagai bentuk. Yeboah-Ofori (2020) menggunakan machine learning untuk memprediksi pergerakan *malware* dan mengklasifikasikannya. Selain itu, Utku, Doğru, dan Akcayol (2018)

menunjukkan bahwa machine learning dapat digunakan pada sistem Android untuk membantu proses deteksi *malware*.

Dalam penelitian ini, metode *decision tree* akan digunakan dengan teknik supervised learning. Teknik ini memungkinkan pembelajaran *decision tree* diawasi dengan sampel *malware* yang telah dilabeli sebelumnya dari proses analisis dinamis. *Decision tree* akan membantu memproses hasil analisis dinamis, dengan tujuan untuk cepat mengkategorisasikan *malware* yang ditemukan dan membuat penelitian lebih efektif.

*Malware* yang diutamakan dalam analisis dinamis adalah *malware* dari serangan *phishing* berbentuk dokumen yang biasanya digunakan dalam perkantoran dan pertukaran data, seperti PDF, DOCX, dan XLSX. Fokus pada *malware* berbentuk dokumen adalah karena banyaknya serangan *phishing* yang dilakukan melalui pengiriman dokumen melalui *email* dan komunikasi digital lainnya.

## 1.2. Rumusan Masalah

Melalui penelitian ini, akan dilakukan upaya untuk memberikan pembuktian dari rumusan masalah berikut:

1. Bagaimana karakteristik dari *malware* yang terlibat dalam serangan *phishing* dapat diidentifikasi dengan menggunakan metode analisis dinamis?
2. Bagaimana karakteristik malware yang didapat dari menganalisis *malware* berbasis dokumen dapat diterapkan dalam *decision tree*?
3. Bagaimana efektivitas model *Decision Tree* dalam menganalisis dan membagi klasifikasi dari *malware* berbentuk dokumen dalam serangan *phishing*?

## 1.3. Tujuan Penelitian

Dari rumusan masalah yang telah disebutkan sebelumnya, tujuan penelitian yang dapat diidentifikasi adalah sebagai berikut:

1. Meningkatkan pemahaman tentang pola perilaku dan fitur-fitur unik dari *malware* serta mengklasifikasi *malware* dari perilaku yang ditemukan berdasarkan metode analisis dinamis.
2. Mengaplikasikan karakteristik yang ditemukan menjadi sebuah data set yang akan digunakan sebagai data latih dalam membuat decision tree.
3. Mengukur keefektifan model *decision tree* dalam membedakan dokumen yang mengandung *malware* berbentuk dokumen dari dokumen-dokumen lainnya menggunakan data uji yang sudah dipersiapkan.

## 1.4. Batasan

Penelitian dilakukan dengan sampel *malware* yang sering digunakan dalam serangan *phishing*, ini dilakukan untuk mengetahui *malware* yang sering kali dipakai dalam serangan *phishing* yang mencakup 80% dari serangan digital seperti yang tertulis dalam laporan "*State of the Phish 2022 Report*" oleh *Proofpoint* (2022) bahwa banyak bisnis diserang melalui metode *phishing*. Dengan memberikan fokus kepada *malware* yang disebar melalui metode serangan terpopuler, diharapkan hasil yang ditemukan dapat membantu meningkatkan keamanan dan menemukan target utama dari *malware* yang disebarkan melalui metode *phishing*.

Batasan penelitian juga mencakup penggunaan sampel *malware* berbentuk dokumen yang dikumpulkan dari sumber-sumber terpercaya yang secara khusus terkait dengan serangan *phishing*. Analisis akan difokuskan pada karakteristik teknis dan perilaku *malware* dalam dokumen-dokumen seperti format dokumen *pdf*, dokumen *docx*, dokumen *xlsx*, dan format dokumen *pptx* yang sering digunakan dalam banyak organisasi. Hal ini bertujuan untuk menyediakan pemahaman yang lebih mendalam tentang metode dan teknik yang digunakan dalam menyebarluaskan *malware* melalui dokumen-dokumen tersebut dalam konteks serangan *phishing*.

Dalam pembagian klasifikasi malware yang akan diteliti dalam penelitian, akan ada 6 klasifikasi *malware* yang akan menjadi fokus klasifikasi dalam penelitian ini, klasifikasi

ini mengikuti taksonomi *malware* modern yang dibuat oleh djenna et al. dalam penelitiannya mengenai cara analisa, deteksi dan mitigasi sebuah serangan *malware*.

## 1.5. Sistematika Penulisan

Penelitian ini terdiri atas 5 bab, ditambah dengan halaman awal yang terdiri atas lembar judul, kata pengantar, lembar pengesahan, daftar isi, daftar tabel, dan daftar gambar. Batang tubuh penelitian ini mengikuti struktur organisasi berikut:

BAB I PENDAHULUAN

Bab ini terdiri dari latar belakang, identifikasi masalah, tujuan, ruang lingkup, serta sistematika penulisan. Dijelaskan ide dan permasalahan yang menjadi dasar dalam penelitian.

BAB II LANDASAN TEORI

Membahas secara singkat landasan teori yang digunakan dalam penelitian dan informasi lain berupa teori, dokumentasi, serta hasil penelitian lain yang mendukung penyelesaian penelitian.

BAB III METODE PENELITIAN

Membahas metode penelitian yang akan dilakukan untuk melaksanakan analisa *malware* secara dinamis, memaparkan alat-alat dan bahan yang akan digunakan selama analisa.

BAB IV HASIL DAN PEMBAHASAN

Memaparkan dan membahas luaran yang diperoleh melalui proses penelitian dalam melaksanakan analisa *malware* secara dinamis menggunakan *decision tree*.

BAB V KESIMPULAN DAN SARAN

Berisi tentang penutupan dan konklusi dari penelitian yang sejauh ini sudah dilakukan selama proses penelitian.