BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan era digital ini menjadikan komputer menjadi suatu kebutuhan fundamental bagi seseorang untuk memudahkan pekerjaan. Manusia berinteraksi dengan perangkat komputer melalui sebuah sistem operasi. Pengguna dapat secara efektif mengerjakan pekerjaannya melalui perangkat lunak yang berupa program aplikasi yang berjalan di atas sistem operasi. Perangkat lunak (*Software*) merupakan data yang diprogram, disimpan, dan diformat dengan fungsi tertentu. *Software* dapat dianalogikan sebagai suatu nyawa dari suatu komputer. Melalui *software* inilah suatu komputer dapat menjalankan perintah sehingga dapat membantu dan memudahkan manusia dalam melakukan pekerjaan. Dibalik fungsionalitas yang sangat bermanfaat tersebut, terdapat sebuah *software* yang diciptakan untuk melakukan tindak kejahatan dan perusakan yang dapat merugikan orang lain. *Software* tersebut dikategorikan sebagai perangkat lunak berbahaya (*Malware*).

Malicious Software atau umumnya dikenal dengan istilah *malware* merupakan perangkat lunak berbahaya yang secara eksplisit dirancang untuk melakukan sebuah aktivitas berbahaya yang berdampak sangat merugikan bagi korbannya. *Malware* merupakan perangkat lunak yang melanggar kerahasiaan dan integritas data dan penyebab kebocoran informasi yang tidak sah [1]. Suatu *malware* yang aktif dapat mengumpulkan data-data sensitif, merusak kinerja sistem dan mendapatkan akses ke perangkat target. Untuk memasukkan sebuah

malware ke perangkat target, *malware* biasanya disusupkan ke dalam jaringan internet sehingga penyebaran dapat terjadi dengan sangat mudah.

Semakin pesatnya perkembangan teknologi tak lepas juga dengan berkembangnya *malware*. Semakin hari *malware* yang muncul semakin canggih dan tekniknya juga semakin sulit dikenali. Badan Siber dan Sandi Negara (BSSN) dalam bssn.go.id mencatat 888.771.736 serangan siber telah terjadi sepanjang bulan Januari-Agustus 2021 dengan jenis serangan lebih banyak dalam bentuk *malware* [2]. Dengan maraknya kasus Covid-19, masyarakat antusias mencari informasi perkembangan dan penanganan pandemi Covid-19. Mekanisme *Work From Home* (WFH) diberlakukan selama pandemi ini untuk keberlangsungan proses kerja dengan meminimalisir penyebaran kasus Covid-19.

Ditengah perjalanan dunia melawan Covid-19, para pelaku kejahatan siber memanfaatkan hal ini untuk mencari keuntungan. Peretas memanfaatkan antusiasme masyarakat ini sebagai pembuka jalan untuk melakukan intrusi yang tidak sah pada infrastruktur teknologi informasi melalui penyebaran *malware* dalam berbagai macam bentuk seperti *virus, ransomware, email spam, trojan, phising, adware* [3], sehingga upaya pencurian data dan insiden siber mudah dilakukan.

Para peneliti kini lebih peduli bagaimana melindungi data sensitif dari serangan *malware* yang semakin hari semakin berkembang. Berbagai macam pendekatan telah diusulkan untuk menanggulangi permasalahan ini. pendekatan tradisional yang telah diusulkan terbukti berhasil dapat mendeteksi *malware*. Pendekatan tradisional ini menganalisis *malware* secara statis dan dinamis. Analisis statis melihat kode atau struktur program tanpa mengeksekusi.

Menganalisis konten *file* dapat mengonfirmasi apakah suatu *file* tergolong berbahaya, memberikan informasi tentang fungsinya, dan juga dapat digunakan untuk menghasilkan serangkaian *signature* sederhana. Analisis dinamis melibatkan evaluasi *behavior* program pada sistem target, dengan mengeksekusi program dan mengamati perubahan perilaku sistem [3]. Analisis dinamis dapat mengamati tindakan program saat dijalankan. Semakin lama *malware* semakin berkembang dengan daya kerusakan lebih masif terhadap sistem, lebih sulit dideteksi dan dengan mudahnya melewati perangkat lunak keamanan seperti *firewall* dan *antivirus* [4]. Berkembangnya *malware* juga memicu pembaruan metode untuk mengatasi permasalahan yang serius ini. Pendekatan tradisional yang terbukti mampu meminimalisir permasalahan ini belum cukup efisien untuk mendeteksi *malware* dengan varian baru. Pendekatan pembelajaran mesin (*machine learning*) juga diusulkan sebagai metode dalam permasalahan ini[5].

Teknik *machine learning* merupakan proses yang mana komputer dapat mendapatkan informasi melalui data masukan, data masukan akan dikenali dan dipelajari struktur polanya, dari pola yang telah dipelajari, akan dibuat sebuah model pembelajaran. Masukan berikutnya dapat dikenali berdasarkan pola yang telah dipelajari. *Machine learning* yang merupakan sub-bidang dari *Artificial intelligence* yang mana bidang yang sedang populer saat ini. Berbagai macam bidang saat ini mengupayakan teknik ini untuk menghasilkan informasi yang lebih berkualitas sebagai acuan untuk menentukan keputusan dan lain sebagainya. Pengembangan model pembelajaran mesin menjadi suatu kebutuhan baru bagi berbagai sektor industri. Dengan mengimplementasikan teknik ini maka permasalahan dapat terselesaikan secara terukur. Berbagai teknik *machine*

learning terbukti berhasil mendeteksi dan mengklasifikasikan *malware* dengan jenis baru secara lebih efisien [6]–[8]. Namun, dengan teknik yang diperbarui pun bukan jadi teknik tersebut tidak mempunyai kelemahan. Berbagai usulan penelitian yang sebelumnya dilakukan dengan pendekatan *machine learning* inipun evaluasi metode yang terukur masih menghasilkan nilai yang tergolong rendah.

Deep learning yang merupakan sub-bidang dari *machine learning* bekerja lebih canggih. *Deep learning* bekerja layaknya sistem otak manusia, sehingga teknik ini memiliki kinerja yang lebih baik [9]. *Neural network* membentuk dasar dari algoritma *deep learning*. *Convolutional Neural Network* (CNN) merupakan metode *machine learning* yang representatif dalam *deep learning*, khususnya untuk masalah pengolahan citra. CNN memberikan manfaat yang signifikan karena secara otomatis memperoleh fitur-fitur penting untuk mengklasifikasikan data dalam proses pembelajaran [10].

Dalam penelitian ini akan diusulkan pendekatan dimana *malware* akan dideteksi dan diklasifikasikan menggunakan teknik *image processing* dengan konsep *deep learning*. Metode yang akan digunakan adalah *Convolutional Neural Network*. *Malware* akan direpresentasikan ke dalam bentuk gambar yang dapat dijadikan sebagai data masukan untuk dipelajari struktur pola dari data gambar sehingga model *training* dapat dibangun. Dari model yang sudah dibangun akan dilakukan pengujian dan evaluasi terkait kemampuan dari model dalam mengklasifikasikan jenis dari *malware*. Diharapkan dari penelitian yang dilakukan akan memperbaiki dan mengatasi masalah terkait hasil pada penelitian sebelumnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, maka dapat dirumuskan permasalahan penelitian sebagai berikut:

1. Apakah metode *convolutional neural network* mampu mendeteksi *malware* dan mengklasifikasikan jenis dari *malware* ?
2. Apakah metode *convolutional neural network* mampu mengenali pola & struktur *malware* dengan varian baru?
3. Apakah parameter *zero padding* pada arsitektur *convolutional neural network* berpengaruh terhadap kenaikan hasil akurasi uji klasifikasi *malware*?

1.3 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Penelitian berfokus terhadap klasifikasi varian dari *malware.*
2. Metode yang digunakan adalah teknik *image processing* yaitu *convolutional neural network.*
3. *Dataset Malimg* [11] yang digunakan berupa kumpulan data *malware* yang sudah tervisualisasikan dalam bentuk citra.
4. Citra yang dihasilkan merupakan representasi dari *binary code* program *malware.*
5. Bahasa pemrograman yang digunakan adalah bahasa *Python.*

1.4 Maksud dan Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Mengimplementasikan teknik *deep learning* menggunakan algoritma *convolutional neural network* dalam deteksi dan klasifikasi *malicious software (malware).*

2. Mengklasifikasikan jenis *malware* ke dalam jenis variannya masing-masing berdasarkan model *training* yang telah dibangun.
3. Mengukur tingkat performa dari metode yang diusulkan dalam deteksi dan klasifikasi *malware*.

1.5 Manfaat Penelitian

Manfaat yang dari penelitian ini adalah ikut serta berkontribusi dalam proses penanggulangan masalah *malware* yang dampaknya sangat merugikan. Penelitian ini juga mengusulkan sebuah pendekatan yang efisien untuk mengatasi permasalahan yang diangkat. Manfaat lain juga ditujukan bagi peneliti lain, yaitu penelitian ini tentunya masih terdapat kekurangan. Oleh sebab itu, terbuka bagi peneliti lain untuk melakukan kajian dan pengembangan dari pendekatan yang diusulkan di masa yang akan datang.

1.6 Metode Penelitian

Dalam pelaksanaan penelitian ini, dilakukan beberapa metode. Metode yang digunakan adalah sebagai berikut :

1.6.1 Metode Pengumpulan Data

Data yang digunakan merupakan *dataset* publik yang berisi kumpulan data *malware* yang sudah dikonversikan ke dalam bentuk gambar. *Dataset* yang digunakan merupakan rujukan dari penelitian yang sebelumnya telah dilakukan. Data digunakan sebagai *inputan* untuk model *training* yang akan dibangun dan kemudian akan dilakukan pengujian terkait performa model dalam mengidentifikasi dan mengklasifikasi data baru.

1.6.2 Metode Uji Coba atau Eksperimen

Dalam penelitian ini akan diimplementasikan teknik pengolahan citra dalam mendeteksi dan mengklasifikasikan *malware*. Data gambar yang representasi *binary code* dari program malware yang diubah kedalam bentuk vektor 8-bit dan diubah menjadi vektor desimal. Vektor desimal kemudian diubah menjadi matriks 2 dimensi yang kemudian divisualisasikan ke dalam bentuk *grayscale image*. Dengan menggunakan algoritma *convolutional neural network* akan dibangun model *training* untuk mengenali dan mengidentifikasi pola dan struktur *malware* yang sebelumnya sudah divisualisasikan dalam bentuk citra sehingga model mampu mengelompokkan *malware* tersebut berdasarkan variannya. Dalam metode ini akan terjadi dua tahapan yaitu tahap *feature extraction* dengan proses konvolusi dan tahap *fully connected* yang mana pada tahap ini proses sistem mengenali & mengidentifikasi struktur dan pola dari data masukan. Dalam pengujian digunakan *software Google Colab* dengan menggunakan bahasa *Python*. Model *convolutional neural network* dibangun menggunakan bantuan *library keras*. Semua proses komputasi dijalankan melalui platform yang tersedia pada *runtime environment google colab*. Model dibangun dengan tujuan dapat mengkategorikan setiap *malware* ke dalam kelasnya masing-masing. Dengan skema model akan belajar & mengenali pola dan struktur dari data gambar masukan, model yang terbangun diharapkan mampu mengenali dan mengidentifikasi dari uji *malware* sehingga dapat diukur bagaimana model yang dibangun apakah baik dalam mengkategorikan *malware* secara tepat.

1.6.3 Metode Analisis

Analisis dilakukan dengan mengukur tingkat performa model yang dibangun apakah mampu mengidentifikasi *malware* dengan benar. Penentuan kerangka dari arsitektur dan parameter yang membangun suatu model apakah akan mempengaruhi performa model. Pengaruh jenis data, dimensi data dan bobot dari data tiap kelas juga dijadikan sebagai acuan tingkat performa dari pengujian model. Model akan diuji untuk mengklasifikasikan *malware* pada data uji untuk mengukur performanya. Melalui pengujian ini dapat diukur sejauh mana performa dari model. Hasil yang didapatkan akan di komparasi dengan hasil pengujian yang sebelumnya didapatkan. Model yang memiliki performa yang terbaik akan diambil sebagai hasil akhir dari penelitian ini.

1.7 Sistematika Penulisan

Sistematika yang merupakan kerangka dan pedoman penulisan skripsi ditujukan mempermudah baca dan mengetahui pembahasan yang ada pada skripsi ini secara menyeluruh. Adapun sistematika penulisannya adalah sebagai berikut :

BAB I PENDAHULUAN

Bab ini terdiri dari latar belakang, rumusan masalah, batasan masalah, maksud & tujuan penelitian, manfaat penelitian, dan sistematika penulisan skripsi.

BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan tentang tinjauan pustaka yang digunakan sebagai referensi dan rujukan dari penelitian sebelumnya. Bab ini juga memuat tentang dasar-dasar teori

dari objek penelitian dan metode pendekatan yang digunakan pada penelitian kali ini.

BAB III METODE PENELITIAN

Bab ini memuat uraian tentang metode yang digunakan pada penelitian ini, uraian tentang bagaimana penelitian dilakukan, serta alur bagaimana pendekatan yang diusulkan akan mengatasi masalah yang diangkat pada penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Bab ini merupakan bagian utama pada penelitian ini, berisi implementasi pendekatan yang diusulkan. Analisis dari hasil yang didapatkan pada pengujian akan dibahas tuntas pada bab ini. Bab ini memuat hasil dari pengujian yang telah diukur dengan berbagai skenario pengujian yang telah dilakukan.

BAB V PENUTUP

Bab ini memuat jawaban dari tujuan penelitian yang dilakukan. Apakah hasil akhir dari rangkaian proses penelitian yang telah dilakukan mampu menjawab tujuan penelitian untuk mengatasi masalah yang diangkat. Bab ini juga memuat kekurangan dari penelitian sebagai saran yang dapat dilakukan kajian dan pengembangan oleh peneliti berikutnya.