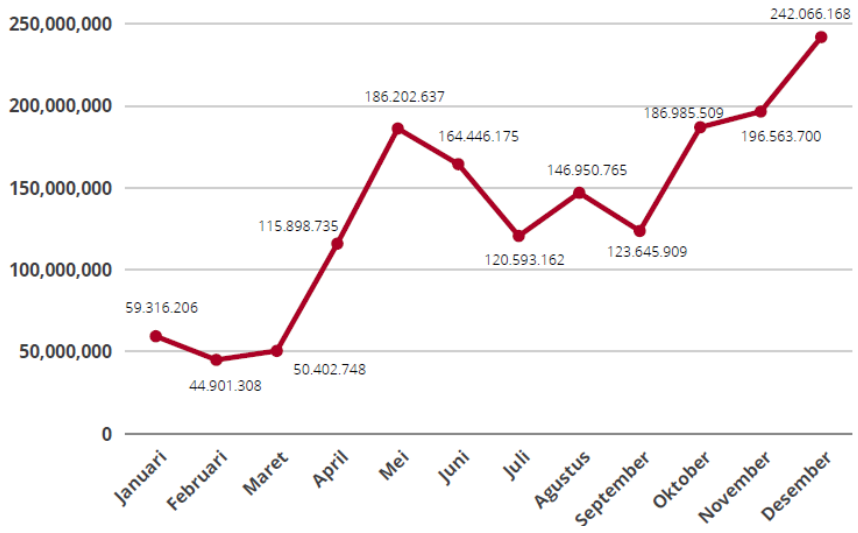
BAB I
PENDAHULUAN

1.1 Latar Belakang

Internet menjadi bagian yang berkembang di kehidupan sosial masyarakat dari hari ke hari. Per April 2022, terdapat lebih dari lima miliar pengguna Internet di seluruh dunia, mewakili 63,1% dari populasi dunia [1]. Internet adalah kumpulan global dari jutaan komputer, jaringan, dan perangkat terkait yang saling terhubung. Inovasi dalam sistem komputer, jaringan, dan perangkat seluler telah meningkatkan penggunaan internet. Dengan banyaknya pengguna internet, sistem komputer dan jaringan juga menjadi lebih rentan terhadap serangan dunia maya [2].

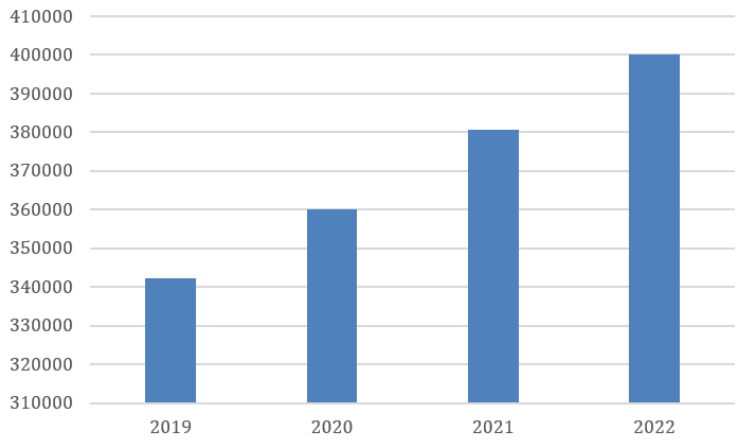
Sistem komputer yang aman dan stabil harus memastikan *confidentiality* (kerahasiaan), *integrity* (integritas), dan *availability* (ketersediaan) data agar tidak terganggu oleh serangan. Integritas dan keamanan sistem komputer terganggu ketika orang atau program yang tidak berwenang memasuki komputer atau jaringan dengan maksud untuk merusak atau mengganggu aliran aktivitas normal [3]. Personil yang tidak berpengalaman atau tidak terlatih, prosedur yang kurang tepat, dan kesalahan konfigurasi merupakan contoh kerentanan dalam membangun sistem jaringan sehingga membuatnya rentan terhadap serangan dunia maya [4].

Berdasarkan laporan tahunan bertajuk "Monitoring Keamanan Siber" untuk tahun 2021 yang dipublikasi di situs resmi Id-SIRTII/CC (*Indonesia Security Incident Response Team on Internet Infrastructure/Coordination Center*) yang berada di bawah Direktorat Operasi Keamanan Siber Badan Siber dan Sandi Negara (BSSN) terungkap bahwa ada lebih dari 1,6 miliar atau tepatnya 1.637.973.022 anomali trafik atau serangan siber (*cyber attack*) yang terjadi di seluruh wilayah Indonesia sepanjang tahun 2021 [5]. Data itu diperoleh dari hasil pemantauan dan identifikasi potensi serangan siber selama 24 jam penuh setiap hari oleh Pusat Operasi Keamanan Siber Nasional BSSN, mulai dari 1 Januari hingga 31 Desember 2021. Gambar 1.1 menampilkan grafik yang menggambarkan jumlah serangan siber yang terjadi di Indonesia sepanjang tahun 2021. Kategori serangan siber terbanyak yang tercatat oleh BSSN adalah *malware* [5].



Gambar 1.1 Grafik jumlah serangan siber yang terjadi di Indonesia sepanjang tahun 2021 [5]

Saat ini, pertumbuhan jumlah dan variasi *malware* yang pesat menjadi salah satu ancaman terbesar bagi keamanan jaringan. *Malware (malicious software)* adalah perangkat lunak jahat yang dirancang untuk melakukan tindakan berbahaya tanpa persetujuan pengguna [6], termasuk akses tidak sah, mencuri informasi sensitif, pencurian lembaga keuangan, merusak sistem komputer, dan memungkinkan eksekusi kode jarak jauh [7]. Penyerang menginfeksi sistem komputer dengan memanfaatkan kerentanan sistem operasi, kerentanan perangkat lunak, atau kecerobohan pengguna [8]. Beberapa contoh kecerobohan pengguna yang dapat menyebabkan sistem komputer terinfeksi *malware* antara lain adalah saat pengguna mengunduh *file* seperti *software*, film, musik, atau *game* populer secara ilegal dan mengklik tautan ke situs web berbahaya yang tersebar di *e-mail* atau postingan jejaring sosial. Berdasarkan laporan dari Kaspersky, terdeteksi sekitar 380.000 *file* berbahaya setiap hari pada tahun 2021. Angka tersebut mengalami peningkatan dalam 10 bulan terakhir tahun 2022, dengan lebih dari 400.000 sampel *malware* baru terdeteksi setiap hari, meningkat sekitar 5% dibandingkan tahun sebelumnya. Peningkatan ini menunjukkan bahwa serangan *malware* terus berkembang, dan mayoritas dari serangan tersebut masih ditargetkan pada sistem operasi Windows [9].



Gambar 1.2 Jumlah harian rata-rata *malware* yang terdeteksi oleh Kaspersky dari 2019 hingga 2022 (1 Januari - 31 Oktober) [9]

Pengembangan teknik deteksi dan analisis *malware* penting dilakukan untuk mengurangi risiko kerentanan keamanan dalam sistem informasi [10]. Proses deteksi *malware* yang semakin kompleks mendorong penggunaan *machine learning* dan *deep learning* untuk mengklasifikasikan *malware*. Beberapa ahli keamanan *cyber* yakin bahwa penggunaan perangkat lunak anti-*malware* berbasis *machine learning* dan *deep learning* akan secara signifikan meningkatkan deteksi jenis *malware* baru serta meningkatkan kinerja mesin pemindaian yang sudah ada [11] [12].

Penerapan *machine learning* dan *deep learning* memberikan keunggulan signifikan dalam bidang deteksi dan analisis *malware*. Teknik *machine learning* konvensional bergantung pada teknik rekayasa fitur dan representasi data yang mengharuskan pemahaman mendalam tentang berbagai domain [13]. Saat ini, *deep learning* yang merupakan model *neural network* yang disempurnakan, berhasil mengungguli metode *machine learning* konvensional dalam banyak tugas, seperti identifikasi pola, pengenalan suara, klasifikasi gambar, dan beragam tugas klasifikasi lainnya. Keunggulan utama dari *deep learning* terletak pada kemampuannya untuk secara otomatis mengekstraksi fitur dari setiap *layer*, karena memiliki kemampuan pembelajaran representasi fitur tingkat tinggi yang diperoleh melalui proses *training* berlapis mulai dari tingkat terendah ke tingkat tertinggi [14].

Penelitian mengenai deteksi dan klasifikasi *malware* sudah banyak dilakukan sebelumnya, tentunya dengan menggunakan metode dan jenis *malware* yang berbeda-beda. Saxe dan Berlin [15] mengemukakan suatu konsep sistem klasifikasi *malware* yang memanfaatkan *neural network* yang memiliki dua lapisan tersembunyi. Lapisan-lapisan tersembunyi tersebut terdiri dari 1024 *Parametric Rectified Linear Units* (*PReLu*), sementara lapisan *output* menggunakan neuron *sigmoid* untuk mengklasifikasikan *malware* atau *benign*. Dalam percobaan ini, *neural network* dilatih menggunakan metode *backpropagation*. Eksperimen ini berhasil mencapai tingkat deteksi sebesar 95%, dengan *false positive rate* sebesar 0,1%, saat diuji pada *dataset* berisi 400.000 sampel. Meskipun demikian, terdapat potensi untuk meningkatkan kinerja metode ini melalui eksplorasi lebih lanjut terhadap fitur-fitur yang digunakan, optimasi *hyperparameter*, serta penggunaan model yang lebih kompleks.

Makandar dan Patrot [16] melakukan eksplorasi dalam klasifikasi sampel *malware* dengan mengubahnya menjadi gambar *grayscale*, kemudian menerapkan analisis berbasis fitur tekstur. Dalam upaya ini, mereka melakukan ekstraksi sebanyak 512 vektor fitur menggunakan *Gabor Wavelet Transform* (*GWT*) dan *Generic Fourier Descriptor* (*GIST*) guna memahami perilaku dari sampel-sampel *malware* tersebut. Dari seluruh vektor fitur tersebut, mereka memilih 320 fitur dua dimensi untuk selanjutnya diterapkan pada metode *Artificial Neural Network (ANN)* dalam proses klasifikasinya. Pemilihan penggunaan *feed-forward artificial neural network* didorong oleh kemampuannya untuk mengidentifikasi pola tersembunyi dalam data kompleks, seperti gambar, suara, dan sinyal, serta memberikan fleksibilitas implementasi. Selama proses *training*, digunakan algoritma *backpropagation*. Meskipun telah berusaha keras, para penulis menghadapi tantangan dalam mengelola vektor fitur berdimensi tinggi secara efektif, yang berdampak pada akurasi suboptimal sebesar 96,35% untuk tugas klasifikasi *malware*.

Babak et al. [17] mengembangkan model *neural network* untuk mengklasifikasikan *file PE*, menggunakan algoritma *Gradient Descent* dengan *decayed learning rate* untuk memperbarui bobot selama proses *training*. Sebelum

proses *training* dan *testing* model *neural network* dilakukan, *dataset* yang digunakan di-*labeling* terlebih dahulu dengan langkah-langkah tertentu, termasuk pengelompokkan sampel, pemeriksaan validitas *PE*, pengecualian *malware* terbungkus, dan pemeriksaan melalui VirusTotal API. Proses ini dilakukan untuk memastikan bahwa setiap sampel dalam *dataset* mendapatkan label yang benar. Penelitian ini menerapkan metode *backpropagation*, di mana setiap *forward pass* diikuti oleh *backpropagation*. Model *neural network* yang dibangun menggunakan fungsi aktivasi *softmax* untuk neuron *output*. Dalam eksperimen ini, digunakan *dataset* yang terdiri dari 4.000 sampel (3.000 *file PE malware* dan 1.000 *benign*). Hasil implementasi model *neural network* ini mencapai metrik kinerja yang mengesankan. Dengan rata-rata akurasi sebesar 97,8%, serta presisi dan *recall* berturut-turut mencapai 97,6% dan 96,6%.

Ahmed et al. [18] menggunakan dua model *neural network* yaitu *feed-forward backpropagation ANN*, dimana dalam penelitian ini, algoritma *backpropagation* digunakan dalam pengembangan model tersebut. Model ini juga melibatkan model *deep neural network* dengan optimasi algoritma *Adam*. Penelitian ini menggunakan *dataset* CTU-13 yang mencakup berbagai skenario serangan *botnet* dan lalu lintas normal. Hasil penelitian menunjukkan bahwa model *deep learning* yang diusulkan dapat mengenali serangan *botnet* dengan akurasi tinggi, mencapai 99,6%. Namun, kelemahan metode ini adalah kompleksitas model yang berpotensi mengalami *overfitting* pada data *training* dan model yang sulit diinterpretasi.

Penelitian sebelumnya telah menunjukkan bahwa *Back Propagation Neural Network* memiliki potensi yang menjanjikan dalam mendeteksi *malware*, karena mampu mengenali pola dan hubungan yang rumit dalam data. Akan tetapi, kinerja *Back Propagation Neural Network* sangat dipengaruhi oleh *hyperparameter* yang mengatur struktur jaringan dan proses pembelajaran. Meskipun penelitian sebelumnya telah mengkaji penggunaan *Back Propagation Neural Network* untuk mengklasifikasikan *malware*, penerapan khusus *Back Propagation Neural Network* dengan optimasi *hyperparameter* menggunakan metode *Grid Search* serta penggunaan regularisasi *dropout* masih relatif belum banyak diteliti. Oleh karena

itu, penelitian ini mengadopsi dua pendekatan berbeda dalam menerapkan model *neural network* untuk mendeteksi *malware*.

Pendekatan pertama melibatkan penerapan *Back Propagation Neural Network* tanpa optimasi *hyperparameter tuning*. Namun, dalam upaya meningkatkan performa model *Back Propagation Neural Network*, penelitian ini mengusulkan pendekatan kedua yang melibatkan penerapan *Back Propagation Neural Network* dengan optimasi *hyperparameter* menggunakan metode *Grid Search* serta penggunaan regularisasi *dropout*. Penggunaan regularisasi *dropout* dalam penelitian ini memiliki peran penting dalam mencegah *overfitting* serta meningkatkan kemampuan model untuk menggeneralisasi data yang belum pernah dilihat sebelumnya. Tujuan utama dari penelitian ini adalah membandingkan kedua pendekatan *neural network* tersebut dalam hal performa dan efektivitasnya dalam mendeteksi *malware*. Penelitian ini diharapkan dapat memberikan kontribusi baru dalam pengembangan metode pendeteksian *malware*. Berdasarkan uraian diatas, penulis memilih untuk menetapkan judul penelitian pada laporan tugas akhir dengan judul "Peningkatan Performa *Back Propagation Neural Network* dengan Optimasi *Hyperparameter* untuk Deteksi *Malware*".

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka permasalahan yang timbul dalam penelitian ini adalah :

1. Bagaimana performa penerapan metode *Back Propagation Neural Network (BPNN)* pada sistem deteksi *malware* tanpa optimasi *hyperparameter tuning*?
2. Bagaimana perbandingan hasil performa antara metode *back propagation* tanpa optimasi *hyperparameter tuning* dan metode *back propagation* dengan optimasi *hyperparameter tuning* menggunakan *Grid Search*?
3. Bagaimana performa *website* deteksi *malware* yang dikembangkan dengan menggunakan model *neural network* berperforma terbaik dalam mengidentifikasi dan membedakan antara *file malware* dan *benign*?

1.3 Batasan Masalah

Agar permasalahan yang dibahas pada penelitian ini tidak keluar dari topik pembahasan, maka penulis membatasi masalah yang akan dibahas adalah mengenai hal-hal sebagai berikut:

1. Penelitian menggunakan *dataset* publik, yaitu *dataset* "*Malware Detection in PE Files*" yang terdiri dari 138.047 data, digunakan untuk mengklasifikasikan *file PE* ke dalam 2 kelas, yaitu *malware* atau *benign*.
2. Metrik evaluasi yang digunakan untuk mengukur performa dan membandingkan model *neural network* adalah akurasi, presisi, *recall*, dan *F1-score*.
3. Metrik evaluasi yang digunakan untuk menguji performa *website* adalah akurasi dalam melakukan prediksi terhadap data *malware* dan *benign*.
4. Informasi mengenai *PE header* dari suatu *executable file* akan diperoleh menggunakan *library pefile*.
5. Sistem deteksi *malware* berbasis web yang dibangun hanya akan memberi peringatan kepada pengguna tanpa adanya tindakan lanjut ketika *malware* terdeteksi dalam *file* berformat **.exe*.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan di atas, maka dapat disimpulkan bahwa tujuan dari penelitian ini adalah:

1. Mengetahui performa penerapan metode *Back Propagation Neural Network (BPNN)* pada sistem deteksi *malware* tanpa optimasi *hyperparameter tuning*.
2. Membandingkan hasil performa antara metode *back propagation* tanpa optimasi *hyperparameter tuning* dan metode *back propagation* dengan optimasi *hyperparameter tuning* menggunakan *Grid Search*.
3. Mengetahui performa *website* deteksi *malware* yang dikembangkan dengan menggunakan model *neural network* berperforma terbaik dalam mengidentifikasi dan membedakan antara *file malware* dan *benign*.

1.5 Manfaat Penelitian

Adapun manfaat yang diperoleh jika tujuan penelitian ini tercapai antara lain :

1. Memberikan kontribusi dalam pemahaman lebih mendalam mengenai penerapan metode *Back Propagation Neural Network (BPNN)* dalam konteks deteksi *malware*.
2. Hasil perbandingan antara metode *back propagation* tanpa optimasi *hyperparameter tuning* dan metode *back propagation* dengan optimasi *hyperparameter tuning* menggunakan *Grid Search* akan memberikan wawasan tentang pengaruh optimasi terhadap performa model *BPNN*.
3. Dapat membantu pemerintah, organisasi, dan masyarakat dalam menangani kasus *cyber attack* khususnya *malware* pada pengaplikasian model deteksi.

1.6 Metodologi Penulisan

Dalam penulisan laporan tugas akhir ini menggunakan beberapa metode, yaitu :

1. Metode Studi Literatur

   Studi Literatur adalah tahapan observasi dan pemahaman konsep, teori dan materi pendukung penelitian. Literatur yang dikaji meliputi *malware*, konsep dasar *neural network*, metode *backpropagation*, serta st*rate*gi optimasi *hyperparameter* dalam konteks deteksi *malware*. Referensi dapat diperoleh dari buku, jurnal ilmiah, artikel, laporan penelitian, *e-book*, dan media lain seperti internet sebagai landasan dalam menyusun laporan.
2. Metode Pengumpulan Data

   Pengumpulan data adalah tahap untuk melakukan pengumpulan data sebagai acuan pengambilan informasi. Data yang dikumpulkan berupa data kumpulan *malware* dan *benign*.
3. Metode Eksperimen

   Pada tahap ini merupakan perancangan skenario dan hasil eksperimen yang dilakukan untuk mengevaluasi kinerja model *Back Propagation Neural Network (BPNN)* dalam deteksi *malware*.

1.7 Sistematika Penulisan

Di dalam pembuatan suatu karya tulis, dibutuhkan suatu sistematika penulisan agar pembaca dapat mempermudah dalam memahami dan membaca isi dari laporan tugas akhir ini. Adapun penulisan laporan tugas akhir ini terdiri atas 4 empat bab, yang dapat dikemukakan sebagai berikut :

BAB I PENDAHULUAN

Bab ini memberikan gambaran secara jelas mengenai latar belakang permasalahan, ruang lingkup masalah, maksud dan tujuan, metodologi penulisan dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini membahas tentang informasi yang bersifat umum dan merupakan teori pendukung pada pembahasan masalah berdasarkan referensi serta penelitian sebelumnya yang berkaitan dengan tugas akhir ini.

BAB III METODOLOGI PENELITIAN

Bab ini membahas mengenai kerangka penelitian, proses perancangan sistem, serta evaluasi model penelitian tugas akhir ini.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil yang dicapai dengan menggunakan metodologi yang telah ditentukan sebelumnya.

BAB V KESIMPULAN DAN SARAN

Bab ini merupakan bagian akhir dari penulisan tugas akhir yang berisi tentang kesimpulan serta saran.

DAFTAR PUSTAKA

LAMPIRAN