BAB I PENDAHULUAN

I.1 Latar Belakang

Keamanan jaringan adalah hal yang penting untuk diperhatikan, terutama pada era teknologi saat ini. Banyaknya organisasi maupun individu yang tidak peduli terhadap masalah keamanan yang dimiliki. Sehingga disaat ketika jaringan mendapatkan serangan dan mengalami kerusakan sistem, disaat itu juga harus mengeluarkan biaya untuk melakukan perbaikan sistem yang dirusak (Triyansyah, 2017). Berdasarkan informasi yang diperoleh dari Pusat Operasi Keamanan Operasi Keamanan Siber Nasional (Pusopskamsinas) Badan Siber dan Sandi Negara (BSSN) mencatat bahwa ada sekitar 88.414.296 serangan telah terjadi sejak 1 Januari hingga 12 April 2020. Jumlah serangan maksimum terjadi pada 12 Maret 2020 mencapai 3.344.470 serangan, kemudian jumlah serangan menurun secara signifikan ketika kebijakan *work form home* (WFH) diterapkan di berbagai tempat. Dari beberapa jenis serangan yang paling umum adalah aktivitas *trojan horse* hingga 56%, diikuti oleh *information gathering* hingga 43%, sedangkan 1% adalah *web application attack.* Untuk menjaga keamanan jaringan perlu diterapkan konsep dasar yang biasa dikenal dengan CIA: kerahasiaan (*confidentiality*) adalah konsep yang membatasi untuk mengakses informasi hanya orang yang tepat, integritas (*integrity*) adalah informasi yang didapat akurat dan tidak ada perubahan, dan ketersediaan (*availability*) adalah konsep bahwa informasi akan selalu tersedia disaat orang yang berwenang membutuhkannya dan dapat diakses dengan cepat (Phintraco, 2018).

Untuk melakukan pecegahan terhadap potensi serangan sudah dikembangkan oleh suatu sistem atau metode yang dikenal dengan Intrusion Detection System (IDS). Intrusion Detection System (IDS) merupakan sebuah aplikasi perangkat lunak atau perangkat keras yang dapat mendeteksi aktivitas yang diduga mencurigakan pada sistem atau jaringan. IDS juga digunakan untuk mendeteksi aktivitas-aktivitas yang dianggap mencurigakan dalam sebuah sistem atau jaringan (Gondohanindijo, 2011). Sedangkan intrusion adalah aktivitas tidaksah atau tidak diinginkan yang mengganggu konfidensialitas, integritas dan atau ketersediaan dari informasi yang terdapat di sebuah sistem. IDS memantau lalu

lintas data di jaringan atau memulihkan data dari file log. IDS akan menganalisa dan dengan algoritma tertentu akan memutuskan apakah akan memberitahu administrator jaringan atau tidak (Sutarti, Pancaro, & Saputra, 2018). IDS sendiri mempunyai dua metode untuk melakukan pendeteksian yaitu *Rule Based* dan *Behavior Based*. Pendeteksian *Behavior Based* dilakukan dengan cara membandingkan aktivitas yang ada sebuah dataset menggunakan sebuah metode untuk proses klasifikasi dan akan menghasilkan sebuah model (Fibrianda & Bhawiyuga, 2018). Dari model yang telah dibuat kemudian dilakukan pengujian dengan menggunakan data testing lalu akan menghasilkan output untuk melihat akurasi apakah lalu lintas yang ada dapat dikategorikan sebagai intrusi atau bukan. Maka dalam hal ini diperlukan sebuah metode yang digunakan dalam melakukan proses klasifikasi untuk mendapatkan akurasi yang cukup akurat.

Dari beberapa penelitian yang telah dilakukan untuk membandingkan metode klasifikasi data serangan jaringan komputer. Penelitian pertama dilakukan oleh Mercury Fluorida Fibrianda dan Adhitya Bhawiyuga (2018). Hasil dari penelitian tersebut menyatakan bahwa kinerja algoritma *Naïve Bayes* lebih baik dibanding dengan kinerja algoritma SVM. Penelitian kedua melakukan perbandingan dengan metode *K-Nearest Neighbor* dan *Decision Tree* oleh Ilham Ramadhan, Parman Sukarno dan Muhammad Arief Nugroho (2019). Hasil dari penelitiannya menyatakan bahwa kinerja algoritma *K-Nearest Neighbor* lebih baik dibandingkan dengan kinerja algoritma *Decision Tree*. Penelitian ketiga melakukan perbandingan nilai akurasi dari metode *Probabilistic Neural Network* dan *Naive Bayes* oleh Tri Muryani (2020). Dalam penelitiannya menyatakan bahwa kinerja dari algoritma *Naïve Bayes* lebih baik dibandingkan dengan algoritma *Probabilistic Neural Network.* Penelitian keempat melakukan pengombinasian dari berbagai algoritma diantaranya Decision Tree, *K-Nearest Neighbor*, *Logistic Regression*, dan *SVM* oleh Vinnia Kemala Putri dan Felix Indra Kurniadi (2018). Hasil dari penelitian menyatakan bahwa kinerja algoritma *Logistic Regression* lebih baik dibandingkan kinerja *K-Nearest Neighbor*, *Decision Tree*, dan *SVM*. Dari beberapa penelitian yang telah disebutkan, bahwa diperlukan adanya perbandingan algoritma lain untuk melakukan indentifikasi nilai akurasi dari masing-masing metode untuk klasifikasi data serangan jaringan

komputer. Selain itu juga, belum adanya yang membahas metode klasifikasi untuk membandingkan nilai akurasi dari metode *K-Nearest Neighbor* dan *Naïve Bayes.*

Berdasarkan penjabaran dari permasalahan yang telah dipaparkan sebelumnya, penulis melakukan penelitian yang berjudul "Analisis Perbandingan Akurasi *K-Nearest Neighbor* dan *Naïve Bayes* untuk Klasifikasi Data Serangan Jaringan Komputer". Untuk itu penelitian ini dilakukan agar dapat menghasilkan nilai akurasi deteksi terhadap serangan jaringan komputer.

I.2 Rumusan Masalah

Berdasarkan analisis latar belakang di atas, maka rumusan masalah yang mendasari untuk penelitian ini adalah:

a. Bagaimana akurasi dari *K-Nearest Neighbor* untuk klasifikasi data pada serangan jaringan komputer?
b. Bagaimana akurasi dari *Naïve Bayes* untuk klasifikasi data pada serangan jaringan komputer?
c. Bagaimana perbandingan akurasi dari *K-Nearest Neighbor* dan *Naïve Bayes* untuk klasifikasi data pada serangan jaringan komputer?

I.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, adapun tujuan dari penelitian ini dilakukan untuk:

a. Hasil akurasi dari *K-Nearest Neighbor* untuk klasifikasi data pada serangan jaringan komputer.
b. Hasil akurasi akurasi dari *Naïve Bayes* untuk klasifikasi data pada serangan jaringan komputer.
c. Perbandingan hasil akurasi dari *K-Nearest Neighbor* dan *Naïve Bayes* untuk klasifikasi data pada serangan jaringan komputer.

I.4 Batasan Penelitian

Batasan dalam penelitian analisis perbandingan akurasi *K-Nearest Neighbor* dan *Naïve Bayes* untuk kelasifikasi data pada serangan jaringan komputer adalah:

a. Menentukan nilai akurasi dari masing-masing algoritma menggunakan *confusion matrix.*
b. Melakukan pengujian dengan menggunakan *confusion matrix* kemudian memvalidasi dengan menggunakan kurva ROC untuk menentukan hasil yang terbaik antara kedua algoritma.

I.5 Manfaat Penelitian

Adapun manfaat dari penelitian dalam memberikan informasi dan hasil penelitian memiliki manfaat yang daintaranya:

Manfaat Teoritis:

1. Menunjukkan hasil akurasi pada dataset ISCX Testbed 14 Juni 2012.
2. Memberikan wawasan dalam perbandingan nilai akurasi dari *K-Nearest Neighbor* dan *Naïve Bayes*. Penelitian ini dapat digunakan sebagai acuan referensi penelitian sejenis di masa mendatang.

Manfaat Praktis:

1. Mengetahui script yang bisa digunakan untuk melakukan analisa akurasi.
2. Mengetahui teknik untuk melakukan validasi menggunakan kurva *Receiver Operating Characteristic.*
3. Mengetahui alat yang dapat memudahkan dalam mendeteksi dari serangan.

I.6 Sistematika Penelitian

Pada penelitian ini disusun dengan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini menjelaskan mengenai latar belakang dari penelitian, rumusan masalah, tujuan dari penelitian, batasan yang dilakukan selama penelitian, manfaat yang didapat dari penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini berisi mengenai literatur dan beberapa teori yang berkaitan dengan permasalahan dan dijadikan sebagai acuan dalam penelitian.

BAB III METODOLOGI PENELITIAN

Pada bab ini berisi mengenai gambaran rinci mengenai langkah-langkah yang dilakukan dalam penelitian, termasuk model konseptual dan penyelesaian masalah yang sistematis.

BAB IV ANALISIS PERANCANGAN

Pada bab ini berisi mencakup mengenai langkah-langkah bagaimana pengumpulan data dilakukan dan implementasi lingkungan pengujian selama klasifikasi dataset dengan menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes*.

BAB V IMPLEMENTASI DAN PENGUJIAN

Pada bab ini memaparkan mengenai uraian tahap pengimplementasian dari metode yang didefinisikan dan juga output dari klasifikasi data yang dilakukan, yaitu *confusion matrix* dan kurva ROC. Kemudian akan dianalisa untuk mencari tahu performansi dari nilai akurasi pada setiap metode.

BAB VI KESIMPULAN DAN SARAN

Pada bab ini berisi mengenai kesimpulan dari hasil penelitian yang dilakukan, serta saran yang dapat digunakan untuk mengembangkan pada penelitian selanjutnya.