BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Pada era sekarang dimana perkembangan teknologi berkembang dengan pesat membuat semakin meningkatnya segala aktifitas yang erat kaitannya akan jaringan, baik kebutuhan akses informasi ataupun data. Keamanan jaringan menjadi suatu hal yang penting seiring dengan kebutuhan akan informasi yang terdapat pada jaringan. Penggunaan jaringan yang semakin besar dengan kurangnya upaya menjaga keamanan jaringan, akan membuka peluang terjadinya suatu ancaman seperti ancaman serangan, bahkan tindakan peretasan jaringan *(hacker)*[1].

Anomali trafik menjadi salah satu bentuk ancaman yang terjadi pada jaringan yang dapat menyerang suatu sistem keamanan jaringan. Menurut Amalia dan Rino [2][3] Anomali trafik merupakan suatu keadaan yang menyebabkan abnormalisasi pada lalu lintas jaringan. Penyebab dari anomali ini bisa saja faktor dari banyaknya pengguna internet atau serangan pada suatu jaringan atau adanya aktivitas-aktivitas dalam jaringan yang menyimpang dari batas normal. Jenis serangan yang dapat mengancam suatu keamanan sistem banyak jenisnya diantaranya yaitu serangan *DoS (Denial of Service), DDoS (Distributed Denial of Service), Brute Force, Sql Injection,* dll. Serangan-serangan tersebut apabila tidak dapat dideteksi akan membawa dampak kerugian terutama bocornya informasi penting ke pihak luar, sehingga mengakibatnya penyalagunaan data [2].

Untuk menghindari hal tersebut dapat dilakukan dengan upaya pendeteksian terhadap suatu tindakan yang mencurigakan yang merupakan awal dari serangan pada sistem yang dilakukan oleh peretas [1]. IDS *(Intrusion Detection System)* merupakan salah satu upaya menjaga keamanan sistem jaringan serta tujuannya yaitu menentukan atau mengidentifikasi anomali yang terjadi, sehingga membantu mengurangi dampak dari serangan [4][5]. Dikarena hal itu diperlukan suatu metode untuk mendeteksi anomali trafik yang ada pada jaringan komputer, salah satu teknik yang dapat digunakan dalam pengidentifikasian yaitu dengan menerapkan algoritma pada *data mining*. *Data Mining* biasa dimanfaatkan dalam pengolahan data, untuk mengetahui pola yang penting atau menarik dari data yang ada pada *database* ataupun *dataset* yang besar [6]. Metode klasifikasi biasa digunakan dalam pengidentifikasian suatu pola data. Ada berbagai macam algoritma yang termasuk dalam algoritma klasifikasi seperti *Naïve Bayes*, C45, KNN, *Random Forest*, dll.

Beberapa penelitian yang telah dilakukan oleh para peneliti sebelumnya terkait penerapan algoritma *data mining* dalam mendeteksi anomali trafik diantaranya : pada penelitian Identifikasi Serangan *Denial Of Sevice (DoS)* Di Jaringan Dengan Algoritma *Decision Tree* C4.5 yang dilakukan oleh Meirza Pratama, Endang Setyati dan F.X.Ferdinandus mendapatkan hasil akurasi sebesar 90,68% algoritma C4.5 unggul dalam mengidentifikasi serangan [7]. Pada penelitian Deteksi Serangan *Botnet* pada Jaringan *Internet of Things* yang dilakukan oleh Mohammad Sani Rafsanjani, Vera Suryani dan Rizka Reza Pahlevi, menggunakan Algoritma *Random Forest* (RF) dengan hasil akurasi

99,27% kinerja *Random Forest* dinyatakan cukup baik dalam melakukan deteksi serangan [8]. Dalam penelitian Komparasi *Performa Tree-Based Classifier* Untuk Deteksi Anomali Pada Data Berdimensi Tinggi dan Tidak Seimbang yang dilakukan oleh Kurniabudi, Abdu Harris dan Veronica, dengan menggunakan kombinasi algoritma *REPTree, J48, Random Tree, Random Forest* dan seleksi fitur *Chi-Square* didapatkan hasil *random tree* dan *random forest* direkomendasikan sebagai sistem deteksi trafik [9].

Dengan adanya berbagai penelitaan tersebut menjadikan algoritma *data mining* penting untuk mendeteksi dan mengidentifikasi jenis serangan atau deteksi intrusi anomali dalam jaringan. Namun belum terdapat penelitian yang melakukan komparasi kinerja algoritma *Naïve Bayes*, C4.5,dan *Random Forest* sehingga belum diketahui metode yang paling akurat dalam mengidentifikasi trafik anomali yang ada pada trafik jaringan komputer. Berdasarkan latar belakang yang telah dijabarkan sebelumnya, melatar belakangi penulis untuk melakukan penelitian dengan judul "**KOMPARASI KINERJA ALGORITMA KLASIFIKASI DATA MINING DALAM MENDETEKSI ANOMALI TRAFIK PADA JARINGAN KOMPUTER**".

1.2 RUMUSAN MASALAH

Adapun rumusan masalah yang akan diselesaikan dalam penelitian ini, sebagai berikut :

1. Bagaimana penerapan algoritma *Naïve Bayes*, C4.5, dan *Random Forest* untuk mengklasifikasi data trafik anomali pada trafik jaringan?
2. Bagaimana tingkat performa klasifikasi algoritma *Naïve Bayes*, C4.5, dan *Random Forest* dalam mendeteksi trafik anomali pada trafik jaringan?

1.3 BATASAN MASALAH

Untuk menghindari terjadinya pembahasan di luar ruang lingkup malasah maka penulis membuat batasan masalah yang akan dilakukan sebagai berikut :

1. *Dataset* yang digunakan berasal dari *dataset* CICIDS2017 yang terdiri dari 170.366 *record* data dan 78 atribut.
2. Serangan yang dideteksi yaitu trafik anomali *Web Attack Brute Force*, *SQL Injection* dan *XSS*.
3. Algoritma *Naïve Bayes*, C4.5, dan *Random Forest* digunakan dalam mengklasifikasi trafik anomali.
4. Seleksi fitur *(Feature Selection)* yang digunakan pada penelitian ini yaitu *Information Gain*.
5. Menggunakan *tools* WEKA sebagai alat bantu analisis.
6. Mode pengujian (validasi) yang digunakan yaitu *Use Training Set* (*full dataset*), 5-*Fold Cross Validation*, dan 10-*Fold Cross Validation.*

7. Menggunakan *Accuracy*, TP (*True Positive Rate*), FP (*False Positive Rate*), *Precission, Recall,* dan ROC/AUC untuk menguji performa algoritma yang dipakai dalam mendeteksi anomali trafik.

1.4 TUJUAN PENELITIAN

Berdasarkan perumusan masalah diatas, berikut tujuan yang dapat dicapai dari penelitian ini yaitu :

1. Membandingkan performa algoritma *Naïve Bayes*, C4.5, dan *Random Forest* dalam mendeteksi serangan atau trafik anomali pada trafik jaringan.
2. Menghitung akurasi deteksi serangan pada trafik jaringan menggunakan algoritma *Naïve Bayes*, C4.5, dan *Random Forest*.

1.5 MANFAAT PENELITIAN

Adapun manfaat yang didapatkan dengan dilakukannya penelitian ini sebagai berikut :

1. Menghasilkan metode deteksi serangan atau trafik anomali dengan performa pengklasifikasian yang baik.
2. Sebagai rujukan bagi peneliti di bidang *Instrussion Detection System.*

1.6 SISTEMATIKA PENULISAN

Untuk memberikan suatu gambaran yang jelas mengenai keseluruhan penulisan ilmiah dibuatlah sistematika penulisan, agar dalam penyusunan penelitian dapat dilakukan secara sistematik membahas topik yang di angkat dan

mengindari terjadinya pembahasan diluar judul penelitian, dapat dilihat sistematika penulisan berikut meliputi :

BAB I : PENDAHULUAN

Bab ini membahas latar belakang, perumusan masalah, pembatasan masalah, tujuan dan manfaat penelitian serta sistematika penulisan.

BAB II : LANDASAN TEORI

Bab ini berisikan teori-teori yang mendukung penelitian serta berisikan konsep-konsep teoritis yang berkaitan dengan pembahasan penelitian, didapatkan dari studi literatur baik dikutip dari buku, jurnal dan lain-lain, serta berisi tinjauan penelitian sejenis.

BAB III : METODOLOGI PENELITIAN

Bab ini berisikan tentang kerangka kerja penelitian, metode yang digunakan, rancangan eksperimen yang akan dilakukan serta *tools* yang akan dipakai dalam penelitian sebagai alat bantu dalam menjawab permasalahan penelitian.

BAB IV : KLASIFIKASI DATA UNTUK DETEKSI ANOMALI

Bab ini berisikan tentang analisa terhadap *dataset* yang dipakai meliputi analisa *dataset*, analisa trafik jaringan, deteksi serangan menggunakan metode yang dipilih, berisi hasil analisa dari uji coba yang telah dilakukan atau hasil deteksi anomali trafik (serangan) yang ada pada *dataset,* dan komparasi dari pengujian performa algoritma yang dipakai.

BAB V : PENUTUP

Bab ini berisikan kesimpulan dan saran dari penelitian yang telah dilakukan untuk pengembangan penelitian lebih lanjut.