

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Sekarang pada kenyataannya, banyak sekali sebuah data yang berukuran besar tidak akurat, tidak komplit dan tidak konsisten. Sebuah data yang tidak berkualitas akan menghasilkan hasil proses yang tidak berkualitas. Yang menyebabkan data tidak akurat, tidak komplit dan tidak konsisten diantaranya adalah kesalahan dari manusia dan *computer error* pada saat memasukan data. Selain itu yang menyebabkan data tidak komplit diantaranya tidak konsisten dalam kaidah penamaan dan tidak konsisten dalam format untuk pengisian. Dengan adanya data yang tidak konsisten akan membuat data yang relevan kemungkinan tidak terekam dan menjadi sulit untuk dimengerti. Apabila data tidak konsisten terekam maka data tersebut akan dihapus secara otomatis.

*Preprocessing* merupakan teknik dari data *mining* untuk mengolah suatu data mentah menjadi data yang berkualitas. Dalam proses *Preprocessing* terdapat beberapa teknik yang digunakan terdiri dari *Data Cleaning*, *Data Integration*, *Data Reduction*, dan *Data Transformation and Data Discretization* [1]. Proses *preprocessing* sendiri akan menghasilkan tujuan diantaranya data tersebut dapat dikelompokkan berdasarkan objek dan attribute. Dengan adanya proses *Preprocessing* maka sebuah kesalahan dalam data akan berhasil dikurangi. Pada penelitian sebelumnya pada tahap *preprocessing*, algoritma yang digunakan adalah *Principal Component Analysis* yang memiliki keuntungan pendeteksian serangan tanpa ada *error* pada pengklasifikasian serangan dan data yang digunakan menggunakan KDD'99 [2][5], sementara untuk *sampling*, algoritma yang digunakan adalah *Stratified Random Sampling* memiliki keuntungan dengan pengambilan sampel secara strata [14][15]. Hasil yang disimpulkan adalah *Principal Component Analysis* salah satu metode *Dimensionality Reduction* untuk mengekstrak fitur dari sebuah dimensional *vector* yang tinggi dari sebuah data [3][4]. Sementara metode *sampling* merupakan salah satu metode untuk mereduksi jumlah data.

Pada Tugas Akhir ini, algoritma yang digunakan untuk *Dimensionality Reduction* adalah *Principal Component Analysis* dan *Sampling* menggunakan *Stratified Random Sampling*. Pada penelitian tugas akhir ini membandingkan dan melakukan penggabungan algoritma *Principal Component Analysis* dan *Stratified Random Sampling* untuk mendapatkan nilai kompleksitas pada masing-masing algoritma. *Raw* data yang dihasilkan berasal dari proses *streaming* simulasi *Network Simulator 3* yang dibuat dengan skema penyerangan. *Raw* data dapat menghasilkan *ouput .tracefile (.tr)*. *File* ini berisikan paket sesuai dengan standar dari NS-3. Setelah menghasilkan *ouputan* lalu akan diolah kembali kedalam proses ke *Sampling* dan *PCA*, sehingga data yang dihasilkan berisi data yang baik, yang nantinya akan diproses kedalam pendeteksian yang menggunakan algoritma *minkowski with k-means*[12]. Setelah diproses kedalam algoritma pendeteksian akan dilakukan pengujian perbandingan kompleksitas. Dalam proses pengujian tersebut akan terlihat nilai kompleksitas pada masing-masing algoritma.

## 1.2 Perumusan Masalah

Pada penelitian Tugas Akhir ini memiliki fokus untuk mengolah *raw* data yang berasal dari proses simulasi pada NS-3 yang *ter-capture* paket datanya[11]. Pada hasil *capture* simulasi tersebut terdapat paket normal dan paket serangan. Hasil *capture* yang didapat akan diolah melalui tahap *preprocessing*. Pada teknik *preprocessing* yang dapat digunakan untuk pengolahan *raw* data adalah *Principal Component Analysis*[2] dan *Stratified Random Sampling*[6]. Sehingga mendapatkan perumusan masalah sebagai berikut :

1. Bagaimana cara menghasilkan *datastream* menggunakan simulasi NS-3.
2. Bagaimana cara melakukan pereduksian data menggunakan metode *Dimensionality Reduction* dan *Sampling*.
3. Bagaimana cara mendapatkan nilai kompleksitas metode *Dimensionality Reduction* dan *Sampling*.

## 1.3 Tujuan

Tujuan dalam penelitian Tugas akhir ini adalah:

1. Membuat sebuah skema jaringan yang terdiri dari beberapa serangan dengan menggunakan simulasi NS-3.
2. Membuat sistem pereduksian pada sebuah *raw data* menggunakan metode *Sampling* dan *Dimensionality Reduction*.
3. Menguji *Time Complexity* antara *Dimensionality Reduction* dan *Sampling*.

#### **1.4 Batasan Masalah**

Batasan masalah yang digunakan dalam penelitian Tugas akhir ini adalah :

1. Data yang akan diolah akan menggunakan data *stream* hasil dari simulasi NS-3.
2. Pada proses *preprocessing* pada metode *Dimensionality Reduction* menggunakan algoritma *Principal Component Analysis* sementara metode *Sampling* menggunakan algoritma *Stratified Random Sampling* dalam proses pereduksian data.
3. Hasil keluaran yang dihasilkan berupa *dataset* baru yang terdiri dari *attribute* dan *variable* baru.
4. Membandingkan kompleksitas berdasarkan waktu pemrosesan algoritma antara *Principal Component Analysis* dan *Stratified Random Sampling*.

#### **1.5 Metodologi Penyelesaian Masalah**

Penelitian ini menggunakan metodologi sebagai berikut :

##### **1. STUDI LITERATUR**

Proses mencari referensi dan materi yang berkaitan dengan proses alur permasalahan metode seperti penggunaan *Principal Component Analysis* dan *Sampling* dalam proses pereduksian data, mempelajari penggunaan NS-3, serta mempelajari beberapa jurnal yang dijadikan acuan dalam Tugas Akhir ini.

##### **2. PENGUMPULAN DATA**

Pada pengumpulan data yang digunakan adalah membentuk skema jaringan yang bersifat *stream* terdiri dari 1 *server*, 2 *router*, 3 *client* serangan, dan 3 *client* normal di NS-3. Hasil output dari simulasi berupa *.tracefile* yang akan di

*preprocessing* dan mengalami pereduksian data menggunakan *Principal Component Analysis* dan *Sampling*.

### 3. PERANCANGAN KEBUTUHAN SISTEM

Sistem yang dibuat yaitu sebuah sistem yang dapat mendeteksi serangan yang dapat mengelola hasil *preprocessing* lalu mereduksi data menggunakan *Principal Component Analysis* dan *Sampling*.

### 4. PENGUJIAN SISTEM

Pada tahap ini akan membangun sebuah sistem penyerangan berupa data *stream* yang menghasilkan *raw data* yang akan diolah kedalam proses *preprocessing*. Hasil dari pengujian sistem ini diantaranya berupa tingkat kompleksitas *Principal Component Analysis* dan *Sampling*.

### 5. ANALISIS HASIL PENGUJIAN

Pada tahap ini akan dilakukan analisa terhadap hasil pengujian berupa nilai kompleksitas dari *Principal Component Analysis* dan *Sampling*.

### 6. PENYUSUNAN LAPORAN TUGAS AKHIR

Pada tahap ini, akan melakukan proses pengumpulan dokumentasi yang diperlukan, penyusunan laporan akhir, format laporan akhir yang sesuai dengan kaidah penulisan dan ketentuan yang telah ditetapkan oleh universitas.

## **1.6 Sistematika Penulisan Tugas Akhir**

Adapun sistematika penulisan pada Tugas Akhir ini adalah :

### **BAB I PENDAHULUAN**

Pada bab ini membahas tentang alasan dilakukannya sebuah penelitian serta identifikasi masalah pada penilitan ini. Bagian-bagian dari bab ini adalah latar belakang masalah, rumusan masalah, tujuan, batasan masalah, metodologi penyelesaian masalah dan sistematika penulisan.

## **BAB II TINJAUAN PUSTAKA**

Pada bab ini menjelaskan rujukan pada penelitian yang pernah dilakukan yang berhubungan dengan topik tugas akhir ini.

## **BAB III PERANCANGAN SISTEM**

Pada bab ini menjelaskan tahap-tahap pengembangan metode yang diteliti dan teknik yang digunakan dalam menyelesaikan masalah yang telah dipaparkan.

## **BAB IV HASIL DAN ANALISIS**

Pada bab ini berisi pemaparan dan analisa hasil yang didapat melalui pengujian kompleksitas dari sistem yang dibuat dan analisis terhadap dari pengujian.

## **BAB V KESIMPULAN DAN SARAN**

Pada bab ini berisi kesimpulan dan saran mengenai pengembangan yang dapat dilakukan oleh peneliti yang lain dalam mengembangkan penilitan selanjutnya.