

ABSTRAK

Pada kenyataannya 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

Pada Tugas Akhir ini, melakukan pengolahan data mentah ke tahap *preprocessing* dengan menggunakan teknik data *reduction* yaitu *sampling* dan *dimensionality reduction*. Tujuannya untuk mereduksi kompleksitas data yang diteliti dan hasil dari data *preprocessing* yang diperoleh dapat diklasifikasikan berdasarkan kebutuhan algoritma yang diteliti. Pada proses *sampling* data yang besar akan diolah menjadi data baru secara acak dari data sample yang ada. Sementara pada proses selanjutnya *dimensionality reduction*, data yang mempunyai *high dimensionality* akan direduksi menjadi *lower dimensionality* sehingga akan mendapatkan output berupa *new feature*.

Data yang akan diteliti berupa *raw data* hasil streaming yang dilakukan oleh NS-3. Data streaming yang dilakukan oleh NS-3 terdiri dari serangan normal dan anomali. Data tersebut akan diolah ke tahap *preprocessing*, sehingga akan memperoleh relevansi fitur trafik baru. Hasil dari penelitian ini memperoleh kompleksitas dari masing-masing algoritma. Dengan hasil kompleksitas tersebut maka kompleksitas skenario 1 lebih baik dengan skenario 2. Dengan adanya penggabungan antara *Sampling* + PCA maka diperoleh nilai *big-O* dengan notasi $O(n,p)$. Dengan n sebagai jumlah data analisis *sampling* dan p sebagai jumlah kolom dari analisis PCA.

Kata kunci : *DDoS, Sampling, Dimensionality Reduction, Time Complexity, NS-3*