1. Pendahuluan

1.1 Latar Belakang

intrusion Detecstion System (IDS) telah dikembangkan sejak tahun 1980-an. Pada tahun 1984 - 1986 Dorothy Denning dan Peter Neumann telah melakukan penelitian tentang model pertama IDS real-time dimana penelitian tersebut diberi nama *Intrusion Detection Expert System* (IDES). Pada awalnya IDES merupakan sebuah sistem cerdas berbasis *rule-based* yang terlatih untuk mendeteksi aktivitas mencurigakan. Sistem yang sama telah ditingkatkan dan dikembangkan untuk membentuk apa yang sekarang diketahui sebagai *Next-Generation Intrusion Detection Expert System* (NIDES). NIDES telah memicu peneliti – peneliti lain untuk melakukan riset awal tentang IDS. Projek – projek seperti *Discovery, Haystack, Multics Intrusion Detection* dan *Alerting System* (MIDAS) adalah projek-projek IDS fase awal yang mendapat pendanaan pemerintah Amerika Serikat dan menjadi acuan pengembangan IDS selanjutnya [1]. Penelitian dan pengembangan IDS meliputi beberapa hal seperti algoritma extraksi ciri, seleksi ciri dan klasifikasi. Pada extraksi ciri terdapat permasalahan yang tidak jauh berbeda dengan klasifikasi data IDS, dimana peneliti melakukan penelitian untuk meningkatkan akurasi IDS [2] [3] [4]. Namun demikian Gopo K. Kumchimanchi [2] Chi-Ho [5] menggunakan extraksi ciri untuk mengurangi *false positive rate* yang tinggi dan meningkatkan *detection rate* yang rendah. Gopo K. Kumchimanchi [2] mengusulkan sebuah teknik extaraksi ciri berbasis *Neural network principal component analysis* (NNPCA) dan *Nonlinier component analysis* (NLCA). Sementara Hai Thanh Nguyen [3] menggunakan *Support Vector Machines* (SVM), *Bayesian Network Classification, Regression Trees dan Session Duration Based Instantiantion Stanislav Ponomarev* [4]. Pada tahun 2005 Chi-Ho [5] menggunakan *K-Means dan Ant Colony Clustering Model* (ACCM) untuk menyelesaikan permasalahan deteksi rate dan *false positive rate.*

Seleksi ciri diimplementasikan pada IDS setelah proses ekstraksi ciri untuk mengurangi waktu proses deteksi dan meningkatkan akurasi, seperti pada Fatemah Amiri [6] Chaouki Khammassi [7]. Yang Li [8] menggunakan seleksi ciri untuk meningkatkan deteksi rate yang rendah pada IDS. Fatemah Amiri [6] mengusulkan algoritma *Least Squares Supoort Vector Machine, Genetic Algorithm dan Logistic Regression* (GA-LR) untuk meningkatkan akurasi pada IDS. Dan Yang Li [8] menggunakan *Random Mutation Hill Climbing* (RMHC) dan *Support Vector Machines* (SVMs) untuk menyelesaikan masalah deteksi rate.

Masalah utama pada klasifikasi data IDS yang ingin di selesaikan oleh para peneliti adalah akurasi Gang Wang [9] Ching-Fong Tsai [10] Björn Waske [11] Sakchi Jaiswal [12], false alarm rate Cheng Xiang [13] Ching-Fong Tsai [10] Chi-Ho Tsang [14], dan utilitas waktu deteksi Akashdeep [15] Gang Wang [9]. Gang Wang [9] menggunakan Artificial Neural Network (ANN) dan Fuzzy Clustering untuk menyelesaikan masalah akurasi dan waktu deteksi. Untuk masalah yang sama Ching-Fong Tsai [10] Björn Waske [11] Sakchi Jaiswal [12] Chi-Ho Tsang [14] memanfaatkan Support Vector Machine (SVM). Serta Björn Waske [11] juga menggunakan teknik Import Vector Machine(IVM) dan Ching-Fong Tsai [10] menggunakan teknik klasifikasi *K-Nearst Neigbor* (KNN) pada masalah diatas. Dipihaklain, Chang Xiang [13] mengusulkan teknik klasifikasi berbasis *Bayesian Clustering* dan *Decision Trees* (DT) untuk menguarangi jumlah *false alarm rate*. Terakhir Akashdeep [15] menggunakan ANN *Classifier* untuk menyelesaikan masalah kompleksitas waktu yang rendah dan penggunaan *resource ultilization* yang tinggi pada tahun 2017. Namun demikian hasil yang mereka dapat menujukkan bahwa waktu deteksi serangan masih lama dan akurasi deteksi masih rendah. Selain hal tersebut mean squared error (MSE) yang didapat juga masih besar.

Kinerja *Learning Vector Quantization* (LVQ) untuk klasifikasi data pada bidang *image processing* sangat menjanjikan. Implementasi untuk mendeteksi pola huruf Jepang menunjukkan bahwa LVQ menghasilkan akurasi sebesar 95%. Pada huruf Hiragana Ricky Brian Purnama [16], sdangkan pada huruf Katakana Amalia Sinta Kurnia [17] mendapatkan akurasi di angka 93.21%. Tetapi pada penelitian lainnya dalam mengidentifikasi manusia Neda Kordjazi [18], dan indentifikasi diagnose Alzheimer's Disease Juan M. Górriz [19] didapatkan hasil akurasi klasifikasi yang baik. Neda Kordjazi mengusulkan algoritma LVQNN untuk meningkatkan *Recognition rate* dan akurasi dalam mengidentifikasi manusia. Sedangakan Juan M. Górriz menggunakan algoritma LVQ-SVM untuk mendiagnosa *Alzheimer's Disease* dengan tingkat akurasi sebesar 90% dan sensitifitas 95%. Namun demikian, penggunaan LVQ untuk klassifikasi data pada IDS masih jarang dilakukan.

Topik dan Batasannya

Tugas akhir ini fokus pada pencarian teknik klasifikasi terbaik untuk IDS menggunakan *anomaly based intrusion detection*. Beberapa yang di analisis adalah *Support Vector Machine* (SVM), Import Vector Machine (IVM), dan K-Nearst Neighbour. Alasan pemilihan teknik-teknik tersebut merupakan teknik yang sering digunakan dalam klasifikasi.

Berikut adalah batasan masalah yang digunakan pada tugas akhir ini:

1. Experimen teknik – teknik klasifikasi IDS dilakukan menggunakan simulasi Matlab-R2016b.
2. Peningkatan akurasi dan waktu deteksi difokuskan pada parameter Alpha dan Konstanta.
3. Data yang digunakan untuk experiment adalah NSL-KDD.

Tujuan

1. Melakukan studi literature terkait metode klasifikasi Intrusion Detection berbasis anomaly detection.
2. Melakukan analisis metode terbaik klasifikasi IDS berbasis anomaly detection.
3. Melakukan analisis parameter – parameter yang memungkinkan peningkatan kinerja akurasi dan waktu deteksi pada metode terbaik yang ditemukan.