TESIS

MODEL DETEKSI SERANGAN SSH-BRUTE FORCE BERDASARKAN DEEP BELIEF NETWORK

Disusun oleh:

Nama : **Constantin Menteng**

NIM : **20.51.1283**

Konsentrasi : **Informatics Technopreneurship**

PROGRAM STUDI S2 TEKNIK INFORMATIKA

PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2023

TESIS

MODEL DETEKSI SERANGAN SSH-BRUTE FORCE BERDASARKAN DEEP BELIEF NETWORK

SSH-BRUTE FORCE ATTACK DETECTION MODEL BASED ON DEEP BELIEF NETWORK

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister

Disusun oleh:

Nama	**: Constantin Menteng**
NIM	**: 20.51.1283**
Konsentrasi	**: Informatics Technopreneurship**

PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023

HALAMAN PENGESAHAN

MODEL DETEKSI SERANGAN SSH-BRUTE FORCE BERDASARKAN DEEP BELIEF NETWORK

SSH-BRUTE FORCE ATTACK DETECTION MODEL BASED ON DEEP BELIEF NETWORK

Dipersiapkan dan Disusun oleh

Constantin Menteng

20.51.1283

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Rabu, 7 Juni 2023

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 7 Juni 2023

Rektor

Prof. Dr. M. Suyanto, M.M.

NIK. 190302001

HALAMAN PERSETUJUAN

MODEL DETEKSI SERANGAN SSH-BRUTE FORCE BERDASARKAN DEEP BELIEF NETWORK

SSH-BRUTE FORCE ATTACK DETECTION MODEL BASED ON DEEP BELIEF NETWORK

Dipersiapkan dan Disusun oleh

Constantin Menteng

20.51.1283

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis

Program Studi S2 Teknik Informatika

Program Pascasarjana Universitas AMIKOM Yogyakarta

pada hari Rabu, 7 Juni 2023

Pembimbing Utama

Arief Setyanto, S.Si., M. T., Ph.D.

NIK. 190302036

Anggota Tim Penguji

Alva Hendi Muh S.T., M.Eng., Ph.D.

NIK. 190302493

Pembimbing Pendamping

Dr. Kumara Ari Yuana, S.T., M.T

NIK. 190302575

Hanif Al Fatta, M.Kom

NIK. 190302096

Arief Setyanto, S.Si., M. T., Ph.D.

NIK. 190302036

Tesis ini telah diterima sebagai salah satu persyaratan

untuk memperoleh gelar Magister Komputer

Yogyakarta, 7 Juni 2023

Direktur Program Pascasarjana

Prof. Dr. Kusrini, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Constantin Menteng**
NIM : **20.51.1283**
Konsentrasi : **Informatics Technopreneurship**

Menyatakan bahwa Tesis dengan judul berikut

Model Deteksi Serangan SSH-Brute Force berdasarkan Deep Belief Network

Dosen Pembimbing Utama : Arief Setyanto, S.Si., M.T., Ph.D.
Dosen Pembimbing Pendamping : Hanif Al Fatta, M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 7 Juni 2023

Yang Menyatakan,

Constantin Menteng

HALAMAN PERSEMBAHAN

Dengan rasa syukur yang mendalam, dengan telah diselesaikannya thesis ini penulis mempersembahkannya kepada:

1. Keluarga besar saya yang telah senantiasa membantu menyelesaikan thesis ini.
2. Segenap *civitas* akademik kampus Universitas Amikom Yogyakarta, staf pengajar, karyawan, dan seluruh teman kampus yang online untuk saling bertemu dan sapanya.
3. Teman-teman saya yang baik dari teman dekat, teman sengkatan maupun teman-teman yang disekitar saya yang telah mendukung saya dan mesupport saya untuk menyelesaikan thesis ini. Hingga dapat terselesaikan thesis ini.

HALAMAN MOTTO

"Rahasia kesuksesan adalah mengetahui yang orang lain tidak ketahui"

Aristotle Onassis

"Jangan pergi mengikuti kemana jalan akan berujung. Buat jalanmu sendiri dan tinggalkanlah jejak."

Ralph Waldo Emerson

"Hanya pendidikan yang bisa menyelamatkan masa depan, tanpa pendidikan indonesia tak mungkin bertahan."

Najwa Shihab

"Nilai akhir dari proses pendidikan, sejatinya terekapitulasi dari keberhasilannya menciptakan perubahan pada dirinya dan lingkungan. Itulah fungsi daripada pendidikan yang sesungguhnya."

Lenang Manggala

"Kemajuan kita sebagai bangsa tidak bisa lebih cepat daripada kemajuan kita dalam pendidikan. Pikiran manusia adalah sumber daya fundamental kita."

John F. Kennedy

KATA PENGANTAR

Puji dan syukur saya panjatkan kepada Allah Swt. atas ridho-Nya saya dapat menyelesaikan penyusunan thesis ini. Adapun judul thesis yang saya ajukan adalah " Model Deteksi Serangan SSH-Brute Force berdasarkan Deep Belief Network"

Thesis ini diajukan untuk memenuhi syarat kelulusan mata kuliah Thesis di Universitas Amikom Yogyakarta. Tidak dapat disangkal bahwa butuh usaha yang keras dalam penyelesaian pengerjaan thesis ini. Namun, karya ini tidak akan selesai tanpa orang-orang tercinta di sekeliling saya yang mendukung dan membantu. Terima kasih saya sampaikan kepada:

- Arief Setyanto, S.Si., M. T., Ph.D. selaku dosen pembimbing Pertama saya
- Hanif Al Fatta, M.Kom selaku dosen pembimbing kedua saya.
- Alva Hendi Muh S.T., M.Eng., Ph.D. selaku dosen penguji pertama saya
- Dr. Kumara Ari Yuana, S.T., M.T selaku dosen penguji kedua saya

Semoga segala kebaikan dan pertolongan semuanya mendapat berkah dari Allah Swt. dan akhirnya saya menyadari bahwa thesis ini masih jauh dari kata sempurna, karena keterbatasan ilmu yang saya miliki. Untuk itu saya dengan kerendahan hati mengharapkan saran dan kritik yang sifatnya membangun dari semua pihak demi membangun laporan penelitian ini.

Yogyakarta, 7 Juni 2023

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

INTISARI

Model Deteksi Serangan SSH-Brute Force berdasarkan Deep Belief Network adalah judul dari penelitian ini. Tujuan penelitian ini agar dapat memberikan beberapa pengujian pada DBN seperti mendeteksi akurasi recall dan presisi klasifikasi yang lebih baik, Dengan menggunakan algoritma ini diharapkan kita sebagai penggunanya bisa mengatasi masalah yang cukup sering terjadi seperti serangan brute force ini didalam akun kita maupun didalam perusahaan. Batasan pada penelitian ini adalah mengevaluasi kinerja model dari DBN. Metode yang dilakukan dalam penelitian ini dalam bentuk simulasi untuk penelitiannya.

DBN tidak dapat dibandingkan dengan CNN karna memiliki hasil yang sangat berbeda jauh untuk pengukuran hasilnya. Akan tetapi lebih DBN cukup baik dari logistic regression dan hampir sama dengan nilai dari k-nearest neighbour untuk hasilnya. Jadi algoritma DBN ini lebih baik untuk dibandingkan seperti antara logistic regression dan k-Nearest Neighbour untuk melakukan suatu penelitian didalamnya. Dan hasil dari penelitian ini paling terbaik adalah CNN untuk dapat mendeteksi suatu serangan malware, ransomware, dan serangan siber lainnya seperti brute Force.

Data dari hasil eksperimen tersebut result DBN itu lebih rendah dari hasil penelitian sebelumnya, akan tetapi dengan mendeteksi suatu serangan menggunakan CNN lebih baik untuk dalam hal mendektsi suatu serangan malware maupun ransomware. Dikarenakan hasil dari eksperimen tersebut menunjukan bahwa tingkat akurasi, recal, precesion, dan f1-score itu lebih baik dari DBN.
Kata kunci: DBN, CNN, Logistic regression, k-Nearest, Brute Force

ABSTRACT

SSH-Brute Force Attack Detection Model based on Deep Belief Network is the title of this research. The purpose of this research is to be able to provide some tests on DBN such as detecting better recall accuracy and classification precision. By using this algorithm, it is hoped that we as users can overcome problems that occur quite often such as brute force attacks in our accounts and within the company. The limitation of this study is to evaluate the performance of the DBN model. The method used in this study is in the form of a simulation for research.

DBN cannot be compared to CNN because it has very different results for measuring the results. However, DBN is better than logistic regression and almost the same as the k-nearest neighbor for the results. So the DBN algorithm is better for comparison, such as between logistic regression and k-Nearest Neighbor to do research in it. And the results of this research are the best for CNN to be able to detect malware, ransomware, and other cyber attacks such as brute force.

Data from the experimental results, the DBN results are lower than the results of previous studies, but by detecting an attack using CNN it is better in terms of detecting a malware or ransomware attack. Because the results of the experiment show that the accuracy, recal, precision, and f1-score are better than DBN.

Keyword: DBN, CNN, Logistic regression, k-Nearest, Brute Force

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Meningkatnya jumlah ancaman berbahaya pada jaringan komputer dan layanan Internet karena sejumlah besar serangan membuat keamanan jaringan berada pada risiko yang tak henti-hentinya. Keamanan komputer adalah medan pertempuran antara penyerang dan pembela. Metode yang sangat umum digunakan oleh defenders adalah enkripsi. Enkripsi digunakan untuk memastikan bahwa data aman, bahkan jika orang lain mendapatkan data. Akan tetapi, enkripsi yang merupakan pembela memiliki musuh penyerang yang dikenal dengan istilah dekripsi, yang merupakan sebuah cara untuk menyerang pembela yang melindungi data melalui berbagai cara.

Penyerang jaringan komputer telah memperoleh keterampilan tingkat lanjut dan mengeksploitasi kerentanan yang tidak diketahui untuk mem-bypass solusi keamanan. Di antara serangan jaringan terkemuka adalah serangan brute force. Serangan brute force menjadi lebih sulit untuk berhasil dideteksi pada tingkat jaringan karena tumbuh di mana-mana jaringan berkecepatan tinggi dan meningkatkan volume dan enkripsi lalu lintas jaringan. Aplikasi serangan *brute force* berjalan melalui semua kemungkinan kombinasi karakter legal secara berurutan sampai mereka menemukan input yang benar. Semakin lama kata sandi, semakin banyak waktu yang biasanya diperlukan untuk menemukan input yang benar. Serangan *brute force* yang paling umum menggunakan kamus kata sandi

yang berisi jutaan kata untuk diuji. Serangan brute force yang berhasil tidak hanya memberi peretas akses ke data, aplikasi, dan sumber daya, tetapi juga dapat berfungsi sebagai titik masuk untuk serangan lebih lanjut.(Wanjau, Wambugu, and Kamau 2021) Beberapa tanda dapat ditafsirkan sebagai indikator serangan brute force. Diantaranya termasuk, beberapa upaya login yang gagal dari alamat IP yang sama; login dengan beberapa upaya nama pengguna dari alamat IP yang sama; login untuk satu akun dari banyak alamat IP yang berbeda; upaya login yang gagal dari nama pengguna dan kata sandi yang berurutan menurut abjad; login dengan URL perujuk dari email seseorang atau klien IRC. Peneliti bermaksud untuk memfokuskan pada 1 macam serangan yaitu serangan Brute-force. Agar penelitian ini dapat berfokus pada 1 permasalahan dan tidak mengacu dari serangan yang lainnya. Dan dataset yang digunakan sama dengan penelitian yang diacukan yaitu CSE-CIC-IDS 2018.

Sejak lahirnya konsep intrusion detection oleh Anderson pada tahun 1980 (Anderson, 1980), teknologi machine learning (ML) seperti neural network (NN), k-nearest neighbor (KNN), support vector ma chine (SVM), dan pohon keputusan (DT) telah menjadi pemain kunci dalam penelitian IDS. Namun, karena peningkatan drastis dalam volume dan kompleksitas dalam data lalu lintas jaringan, IDS berbasis ML tradisional dengan struktur dangkal tidak cocok di era IoT dengan miliaran perangkat. Dengan demikian, teknik deep learning (DL) telah diterapkan pada arsitektur NN konvensional dengan nama deep neural network (DNN). Model DNN utama adalah deep belief network (DBN), stacked autoencoder (SAE), convolutional neural network (CNN), dan recurrent neural network (RNN). DBN

adalah DNN yang terdiri dari beberapa mesin Boltzmann (RBM) terbatas yang dilatih dengan cara tanpa pengawasan dan disesuaikan dengan algoritme propagasi balik. DBN adalah teknologi yang paling penting dan paling sering digunakan dalam model IDS yang canggih.(Sohn 2021)

Secure Shell (SSH) adalah salah satu protokol komunikasi paling populer di Internet yang banyak digunakan oleh pengembang, webmaster, dan administrator sistem. SSH memungkinkan seseorang untuk mendapatkan akses jarak jauh ke layanan cloud baru atau kotak khusus hanya dalam hitungan detik menggunakan saluran komunikasi terenkripsi.(Wanjau, Wambugu, and Kamau 2021) Serangan paksa SSH-Brute mencoba untuk mendapatkan akses ke mesin jarak jauh dengan melakukan upaya otentikasi, secara sistematis memeriksa semua kemungkinan kata sandi sampai kata sandi yang benar ditemukan pada protokol Secure Shell. layanan seperti Yahoo atau serangan phishing, mengingat penggunaan kembali kata sandi di seluruh akun tetap merajalela. Alibaba, salah satu perusahaan pengecer dan e-commerce terbesar di dunia mengalami serangan brutal besar-besaran di situs e-commerce-nya, perusahaan *TaoBao* Menggunakan database sekitar 99 juta nama pengguna dan kata sandi, penyerang berhasil mengkompromikan sekitar 21 juta akun. Biasanya, penyerang dapat menggunakan aplikasi dan skrip sebagai alat *brute force*. Alat-alat ini mencoba berbagai kombinasi kata sandi untuk melewati proses otentikasi. Jika host terpapar langsung ke Internet atau *Wide Area Network* (WAN) dan layanan SSH berjalan di host, itu menjadi subjek serangan brute force konstan yang dilakukan oleh skrip otomatis seperti hydra. Dalam kasus lain, penyerang mencoba mengakses aplikasi berbasis

web dengan mencari ID sesi yang tepat. Biasanya, kata sandi yang dipilih manusia bersifat lemah karena pengguna cenderung memilih kata sandi sederhana yang lebih mudah diingat. Terkadang, mereka tidak mengubah kata sandi default mesin atau hanya menggunakan nama pengguna sebagai kata sandi. Hal ini membuat mesin tersebut rentan terhadap serangan *brute force* yang berhasil.

IoT adalah semacam jaringan yang menghubungkan apa pun dengan Internet tergantung pada protokol tertentu melalui perangkat penginderaan data yang mengarah ke berbagi data dan pertukaran dan memungkinkan identifikasi cerdas, pelacakan, penentuan posisi, administrasi, dan monitoring. Definisi reguler IoT adalah sebagai jaringan objek fisik. Internet bukan hanya jaringan PC, namun; itu telah berkembang menjadi jaringan perangkat dari berbagai jenis dan ukuran, peralatan rumah tangga, ponsel pintar, kendaraan, mainan, kamera, toko obat, kerangka kerja modern, orang, hewan, dan struktur, yang semuanya terkait, masing-masing berbagi dan mengkomunikasikan data tergantung pada protokol yang ditentukan.

Secara komprehensif, IDS diklasifikasikan menjadi tiga teknik: deteksi penyalahgunaan, deteksi anomali, dan hibrida. Metode identifikasi penyalahgunaan menggunakan tanda tangan yang telah ditentukan dari tindakan ganas untuk mendeteksi intrusi. Oleh karena itu, mereka digunakan untuk mengidentifikasi serangan known. Metode deteksi anomali mencirikan pola khas dan mendeteksi tindakan berbahaya tergantung pada perbedaannya dari pola biasa. Dengan cara ini, teknik identifikasi berbasis anomali memiliki kemampuan untuk mengidentifikasi serangan zero-day. Metode hybrid mengeksploitasi teknik identifikasi anomali dan

anomali. Dengan mengurangi positif palsu dari serangan yang tidak diketahui, metodologi hibrida menargetkan memperluas tingkat identifikasi intrusi yang diketahui.(Manimurugan et al. 2020)

Deep Belief Networks merupakan model deep learning yang memanfaatkan tumpukan/stack *Restricted Boltzmann Machines* (RBM) atau kadangkala *Autoencoders*. *Autoencoder* adalah model *neural network* yang memiliki input dan output yang sama. *Autoencoder* mempelajari data input dan berusaha untuk melakukan rekonstruksi terhadap data input tersebut.

Penggunaan RBM yang signifikan kemungkinan akan terjadi, ada kelangkaan data berlabel, dan RPM dan encoder otomatis dapat dilatih sebelumnya pada data yang tidak berlabel dan disesuaikan pada sejumlah kecil data berlabel.

Deep belief network dapat digunakan untuk menganalisis model serangan SSH Brute force dengan cara algoritma pelatihan layer-wise digunakan untuk melatih DBN satu lapisan pada satu waktu. Sehingga bisa dilihat seberapa besar akurasi IDN dapat mendeteksi serangan brute force pada sistem ataupun website.

Peneliti bermaksud menggunakan DBN untuk mengetahui kinerja deep learning yang sudah dianggap lama/kuno untuk mengetahui apakah kinerja dari deep belief network tersebut dapat mengatasi suatu serangan cyber security terutama pada Brute Force yang jadi titik acuan untuk diteliti. Tentu saja, peneliti akan membandingkan dengan kinerja pada peneliti sebelumnya dengan membandingkan hasil dari kinerja mendeteksi suatu serangan khususnya di burte-force.

1.2. Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah sebagai berikut:

a. Melakukan evaluasi hasil *accuracy, recall, precission, dan f1-score* pada *DBN*
b. Parameter apa yang mempengaruhi kinerja *deep belief network*

1.3. Batasan Masalah

Bagian ini memuat penjelasan tentang:

a. Mengevaluasikan kinerja model yang dihasilkan dari *Deep Belief Network*
b. Parameter *Deep Belief Network* ini hanya dilakukan perhitungan presisi dan akurasinya.
c. Penelitian ini menggunakan *dataset* sehingga tidak melakukan penelitian secara langsung hanya menggunakan proses perhitungan algoritma *deep belief network* yang berupa *intrusion detection system*.
d. Dataset yang digunakan adalah IDS 2018 intrusion CSVs (CSV-CIC-IDS2018) Membandingkan paper wanjau untuk hasil yang lebih baik untuk mendeteksi serangan Brute Force.

1.4. Tujuan Penelitian

Bagian ini memuat penjelasan secara spesifik:

a. Solusi pada penelitian saya dapat memberikan beberapa pengujian pada DBN seperti mendeteksi akurasi *recall* dan presisi klasifikasi yang lebih baik.

b. Dengan menggunakan algoritma ini diharapkan kita sebagai penggunanya bisa mengatasi masalah yang cukup sering terjadi seperti serangan *brute force* ini didalam akun kita maupun didalam perusahaan.

Tujuan untuk penelitian agar kita dapat mengetahui adanya serangan *brute force* pada akun kita ataupun perusahaan yang kita miliki.

1.5. Manfaat Penelitian

Bagian ini memuat penjelasan tentang:

a. Manfaat dari thesis ini kita dapat memperkecil pelaku yang menyerang akun atau maupun *ecomerce* yang kita miliki.

b. Kegunaan dari menggunakan aplikasi tersebut kita dapat mengetahui jika di akun kita terdapat hacker yang ingin mencoba membobol *password* dan email yang kita gunakan mau itu pribadi atau usaha dan perusahaan yang kita miliki.

c. Manfaat buat penggunanya akun mereka akan terasa aman, dan jika kalau ada perentas kita dapat mengetahuinya lebih cepat.

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Di dalam penelitian ini, Scholarworks (2019) yang berjudul "Utilizing Machine Learning Classifiers to Identify SSH Brute Force Attacks Utilizing Machine Learning Classifiers to Identify SSH Brute Force Attacks Advisor : James Devenick" ini mengeksplorasi penggunaan algoritma pembelajaran mesin dalam mendeteksi dan mencegah serangan semacam *brute force SSH* sebagai Alternatif dari *Firewall* teknik yang digunakan saat ini. Mereka menggunakan tiga pengklasifikasi yang berbeda *naive bayes* , *decision tree* , KNN. Pada kumpulan data yang tersedia untuk umum dari aliran jaringan berlabel untuk mencoba dan mengklasifikasikan aliran jaringan yang tidak dikenal kedalam kategori *brute force* jinak dan SSH. Hasil mereka menunjukan bahwa pembelajaran mesin sangat cocok untuk Mengindefikasi brute force SSH, dengan semua pengklasifikasi memiliki skor akurasi lebih dari 85% dalam klasifikasi data pengujian.

Di dalam penelitian ini, Wanjau, Wambugu, and Kamau (2021) yang berjudul "SSH-Brute Force Attack Detection Model based on Deep Learning" ini mengusulkan mekanisme yang efesien untuk deteksi serangan Jaringan SSH Brute Force berdasarkan algoritma supervised deep learning algorithm, Convolutional Neural Network. Performa model dibandingkan dengan hasil eksperimen dari algoritma *machine learning* klasik antara lain *Naive Bayes, Logistic Regression, Decision Tree, k-Nearest Neighbour (KNN), dan Support Vector Machine (SVM).*

Empat metrik standar yaitu, *Accuracy, Precision, recall, F-measure* digunakan. Hasil dari penelitian menunjukkan bahwa model berbasis CNN lebih unggul dari Metode pembelajaran mesin tradisional dengan akurasi 94,3% tingkat presisi 92,5% tingkat recall 97,8% dan F1- score 91,8% dalam hal kemampuan mendeteksi SSH Brute Force.

Di dalam penelitian ini, Ferrag et al. (2019) yang berjudul "Deep Learning Techniques for Cyber Security Intrusion Detection : A Detailed Analysis" mereka menyajikan analisis rinci teknik pembelajaran mendalam untuk deteksi intrusi. secara khusus, kami menganalisis tujuh *deep learning models, including, deep neural networks, recurrent neural networks, convolutional neural networks, restricted Boltzmann machine, deep belief networks, deep Boltzmann machines, and deep autoencoders*. dan autoencoder yang dalam. Untuk setiap model pembelajaran mendalam, kami mempelajari kinerja model dalam klasifikasi biner dan klasifikasi multiclass. Kami menggunakan set data CSE-CIC-IDS 2018 dan TensorFlow sistem sebagai dataset benchmark dan perpustakaan perangkat lunak dalam percobaan deteksi intrusi.

Di dalam penelitian ini, Ismail et al. (2022) yang berjudul "A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks", menggunakan pendekatan pembelajaran mesin untuk klasifikasi dan prediksi jenis serangan DDoS. Untuk tujuan ini, digunakan algoritma klasifikasi Random Forest dan XGBoost. Untuk mengakses penelitian mengusulkan kerangka kerja lengkap untuk prediksi serangan DDoS. Untuk pekerjaan yang diusulkan, UNWS-np-15

dataset diekstraksi dari repositori GitHub dan Python digunakan sebagai simulator. Setelah menerapkan model pembelajaran mesin, kami menghasilkan matriks kebingungan untuk mengidentifikasi kinerja model. Pada klasifikasi pertama, hasil menunjukkan bahwa *Precision (PR) dan Recall (RE)* adalah 89% untuk Algoritma Hutan Acak. Akurasi rata-rata (AC) dari model yang kami usulkan adalah 89% yang luar biasa dan cukup baik. Pada klasifikasi kedua, hasilnya menunjukkan bahwa *Precision (PR) dan Recall (RE)* sekitar 90% untuk algoritma XGBoost.

Di dalam penelitian ini, (Manimurugan et al. (2020) yang berjudul "Effective attack detection in internet of medical things smart environment using a deep belief neural network", mereka mengusulkan metode berbasis *deep learning* algoritma *Deep Belief Network* (DBN) model untuk sistem deteksi intrusi. Mengenai serangan dan deteksi anomali, CICIDS Dataset 2017 digunakan untuk analisis kinerja model IDS saat ini. Metode yang diusulkan menghasilkan hasil yang lebih baik di semua parameter dalam kaitannya dengan akurasi, daya ingat, presisi, skor F1, dan deteksi kecepatan. Metode yang diusulkan telah mencapai akurasi 99,37% untuk kelas normal, 97,93% untuk kelas Botnet, 97,71% untuk kelas *Brute Force*, 96,67% untuk kelas Dos/DDoS, 96,37% untuk kelas Infiltrasi, 97,71% untuk kelas Port can dan 98,37% untuk serangan Web, dan hasil ini dibandingkan dengan berbagai pengklasifikasi seperti yang ditunjukkan pada hasil.

Di dalam penelitian ini, Highnam et al. (2021) yang berjudul "BETH Dataset: Real Cybersecurity Data for Unsupervised Anomaly Detection Research"

menyajikan dataset keamanan siber BETH untuk deteksi anomali dan analisis di luar distribusi Dengan "anomali" nyata yang dikumpulkan menggunakan sistem pelacakan, dataset mereka berisi lebih dari delapan juta titik data melacak 23 host. Setiap tuan rumah telah menangkap aktivitas jinak dan, paling banyak, satu menyerang, memungkinkan analisis perilaku yang lebih bersih. Di Selain menjadi salah satu yang paling modern dan eks kumpulan data keamanan siber yang intensif tersedia, BETH en memungkinkan pengembangan algoritma deteksi anomali ritma di dunia nyata yang terstruktur secara heterogen data, dengan aplikasi hilir yang jelas.

Didalam Penelitian ini, Sohn (2021) yang berjudul "Deep belief network based intrusion detection techniques: A survey" mereka mengusulkan memberikan tinjauan lengkap model IDS berbasis DBN dari masa lalu hingga sekarang dan juga membantu pembaca memahami arsitektur dasar dari model yang diusulkan, kami memulai makalah dengan gambaran umum tentang kumpulan data dan metrik kinerja yang digunakan dalam komunitas penelitian deteksi intrusi. Kumpulan data yang diperkenalkan dalam makalah ini adalah KDD Cup 99, NSL-KDD, UNSW-NB15, dan ADFA yang merupakan kumpulan data standar yang tersedia untuk umum. Di antara banyak metrik kinerja yang digunakan dalam analisis model deteksi intrusi, presisi, daya ingat, akurasi, dan karakteristik operasi penerima dijelaskan secara rinci. Sebelumnya, tinjauan dari berbagai karya pada model IDS berbasis DBN, deskripsi singkat tentang DBN dengan tumpukan RBM disajikan. Kerangka umum model DBN-IDS dipelajari berdasarkan 14 karya penelitian penting, dimulai dengan karya penelitian yang disajikan oleh Fiore et al.

Didalam penelitian ini, Roder et al (2021) yang berjudul "Deep belief network based intrusion detection techniques: A survey" Karya ini mengusulkan pendekatan baru untuk meningkatkan kinerja "Residual Deep Belief Networks", yang dikenal sebagai Res DBN, dengan Memperkenalkan koneksi berbobot. Jaringan seperti itu mempertimbangkan satu set hyperparameter untuk mengontrol dukungan dari informasi sisa (yaitu, penguatan) dan data asli sebagai input baru untuk lapisan tersembunyi berikutnya. Selanjutnya, kami juga mengusulkan untuk menyempurnakan bobot tersebut melalui optimasi metaheuristik. Untuk tujuan tersebut, lima optimasi metaheuristik dan dua teknik non-metaheuristik dipertimbangkan.

Didalam penelitian ini, Guo (2021) yang berjudul "ScienceDirect ScienceDirect Application Application of of data data fusion fusion based based on on deep deep belief belief network network in in air air quality quality monitoring monitoring" mereka bertujuan pada kurangnya data pemantauan atmosfer dalam pemantauan kualitas udara, makalah ini mengadopsi metode penggunaan model DBN untuk memadukan data multi-sumber untuk melengkapi data pemantauan yang hilang. Setelah verifikasi eksperimen, dapat dipahami dari hasil perbandingan eksperimen bahwa dibandingkan dengan metode menggunakan BP neural network untuk memadukan data multi-sumber, metode berdasarkan model DBN memiliki kesalahan yang lebih kecil dengan nilai pemantauan yang sebenarnya.

2.2. Keaslian Penelitian

Tabel 2.1. Matriks literatur review dan posisi penelitian

Model Deteksi Serangan SSH-Brute Force berdasarkan Deep Belief Network

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	SSH-Brute Force Attack Detection Model based on Deep Learning	Stephen Kahara Wanjau, Geoffrey Mariga Wambugu, Gabriel Ndung'u Kamau, Journal, 2021	Tujuan makalah ini mengusulkan mekanisme yang efisien untuk deteksi serangan jaringan SSH-Brute force berdasarkan algoritma pembelajaran mendalam yang diawasi, Convolutional Neural Network. Performa model dibandingkan dengan hasil eksperimen dari 5 algoritma machine learning klasik antara lain Naive Bayes, Logistic Regression, Decision Tree, k-Nearest Neighbour, dan Support Vector Machine. Empat metrik standar yaitu, Accuracy, Precision, Recall, dan F-measure digunakan.	Dalam penelitian ini, mereka mengusulkan pendekatan untuk deteksi serangan jaringan SSH-Brute force berdasarkan algoritma pembelajaran mendalam yang diawasi. Model berbasis CNN dirancang dan diuji dengan set data CIC-IDS 2018 yang telah diproses sebelumnya untuk eksperimen kami. Data mentah diubah menjadi gambar dan kemudian digunakan untuk pelatihan dan pengujian model.	eksperimen lebih lanjut dapat dilakukan dengan algoritma pembelajaran mendalam lainnya seperti Deep Belief Network (DBN), Generative Adversarial Network (GAN) dan hasilnya dibandingkan dengan model kami. Selain itu, model kami dapat diuji pada dataset benchmark yang berbeda seperti dataset ISCX IDS 2012.	Penelitian yang akan dilakukan menggunakan deep belief network untuk menganalisis model serangan brute force.

Tabel 2.1. Lanjutan

No		Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
2		Utilizing Machine Learning Classifiers to Identify SSH Brute Force Attacks	Dmytro Shmagin, James Deverick, William & Mary, Theses, 2019	Tujuan makalah ini mengeksplorasi penggunaan Algoritma pembelajaran mesin dalam mendeteksi dan mencegah serangan semacam ini sebagai alternatif dari firewall teknik yang digunakan saat ini. Kami menggunakan tiga pengklasifikasi yang berbeda - nave Bayes, K-terdekat tetangga, dan pohon keputusan - pada kumpulan data yang tersedia untuk umum dari aliran jaringan berlabel untuk mencoba dan mengklasifikasikan aliran jaringan yang tidak dikenal ke dalam kategori brute force jinak dan SSH.	Dari pengklasifikasi yang diuji, mereka menemukan bahwa pengklasifikasi nave Bayes adalah yang kuat baseline tetapi secara konsisten berkinerja lebih buruk daripada dua pengklasifikasi lainnya yang diuji, menghasilkan skor F1 rata-rata antara 0,85 dan 0,92. Skor ini masih cukup tinggi dan tunjukkan bahwa bahkan model sederhana pun dapat mengklasifikasikan serangan brute force SSH dengan tingkat keberhasilan yang besar.	-	Penelitian yang dilakukan sama yaitu mengidentifikasi serangan brute force, tetapi metode yang digunakan menggunakan deep belief network

Tabel 2.1. Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
3	Hybrid Deep-Learning Model to Detect Botnet Attacks over Internet of Things Environments	Mohammed Y. Alzahrani, Alwi M Bamhdi, Article, 2021	Dalam penelitian ini, kami mengusulkan sistem yang kuat khusus untuk membantu mendeteksi serangan botnet perangkat IoT. Ini dilakukan oleh secara inovatif menggabungkan model jaringan saraf convolutional dengan memori jangka pendek yang panjang mekanisme algoritma untuk mendeteksi dua serangan IoT yang umum dan serius (BASHLITE dan Mirai) pada empat jenis kamera keamanan. Kumpulan data, yang berisi paket jaringan berbahaya normal, adalah dikumpulkan dari perangkat kamera yang terhubung dengan lab secara real-time di lingkungan IoT.	Tujuan utama pengembangan sistem ini adalah untuk secara cerdas mendeteksi ancaman platform IoT yang serius. Sistem dapat membantu mendeteksi serangan botnet tidak dikenal yang mengancam jaringan IoT.	-	Penelitian yang dilakukan menggunakan metode deep belief network untuk mendeteksi serangan brute force

Tabel 2.1. Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
4	Detection of Username Enumeration Attack on SSH Protocol: Machine Learning Approach	Abel Z. Agghey 1,* , Lunodzo J. Mwinuka 2, Sanket M. Pandhare 3 , Mussa A. Dida 1 and Jema D. Ndibwile 4, Article, 2021	Dalam makalah ini, kami menyelidiki deteksi serangan enumerasi nama pengguna pada protokol SSH dengan menggunakan pengklasifikasi pembelajaran mesin. Kami menerapkan empat pengklasifikasi asimetris pada kumpulan data yang kami hasilkan dikumpulkan dari jaringan lingkungan tertutup untuk membangun berbasis pembelajaran mesin model untuk deteksi serangan. Penggunaan beberapa pembelajar mesin menawarkan spektrum investigasi yang lebih luas kemampuan pengklasifikasi dalam deteksi serangan.	Temuan kami menyiratkan bahwa, termasuk sumber dan port tujuan sebagai fitur input menghasilkan beberapa peningkatan kinerja tanpa mengorbankan kekuatan komputasi. Namun, peningkatan kinerja bervariasi dari pengklasifikasi ke pengklasifikasi berdasarkan sifatnya. Naïve Bayes memiliki peningkatan yang signifikan dari kinerja saat menyertakan informasi port. Fitur Naïve Bayes sepenuhnya tidak tergantung, karenanya, termasuk informasi port menghasilkan peningkatan kinerja yang signifikan.	Dalam pekerjaan di masa depan, kami bertujuan mengumpulkan data dalam jaringan produksi-lingkungan dan mengevaluasi bagaimana model yang dikembangkan akan tampil pada set data langsung dunia nyata. Pembelajaran mendalam teknik juga dapat dimasukkan di masa depan untuk mendeteksi serangan enumerasi nama pengguna.	Penelitian yang dilakukan untuk mendeteksi serangan brute force dengan menggunakan metode deep belief network

Tabel 2.1. Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
5	Effective Attack Detection in Internet of Medical Things Smart Environment Using a Deep Belief Neural Network	S. MANIMURUGAN 1, (Member, IEEE), SAAD AL-MUTAIRI1, (Member, IEEE), MAJED MOHAMMED ABOROKBAH1, NAVEEN CHILAMKURTI 2, (Senior Member, IEEE), SUBRAMANIAM GANESAN3, (Senior Member, IEEE), AND RIZWAN PATAN 4, Journal, 2020	Tujuan dari peneliti Masalah privasi dan keamanan muncul karena berbagai kemungkinan serangan yang disebabkan oleh penyusup. Jadi, disana adalah kebutuhan penting untuk mengembangkan sistem deteksi intrusi untuk serangan dan identifikasi anomali di IoT sistem.	Dalam makalah ini kami membahas dataset secara rinci untuk evaluasi kinerja. DoS/DDoS, Botnet, Brute Force, Web Attack, Infiltration, dan PortScan adalah jenis serangan yang ada dalam dataset ini yang dapat menyebabkan kegagalan sistem IoT. Parameter evaluasi yang digunakan dalam analisis adalah akurasi, penarikan, presisi, tingkat deteksi, dan skor F1. Model yang diusulkan memperoleh hasil yang lebih baik dalam hal semua parameter dibandingkan dengan teknik yang ada.	Di masa depan, IDS yang diusulkan dapat diperluas untuk mendeteksi jenis serangan lain terhadap sistem IoT, dan berbagai dataset deteksi intrusi. Selain itu, metode yang diusulkan ini dapat digunakan tidak hanya dalam deteksi intrusi, tetapi juga dalam klasifikasi dan pengakuan.	Dalam penelitian ini, dataset yang digunakan adalah BETH dataset dengan menggunakan metode deep belief network untuk penelitiannya.

Tabel 2.1. Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
6	Deep Learning Techniques for Cyber Security Intrusion Detection : A Detailed Analysis	Mohamed Amine Ferrag, Leandros Maglaras, Helge Janicke, Richard Smith, Article, 2019	Tujuannya menyajikan analisis rinci teknik pembelajaran mendalam untuk deteksi intrusi. Secara khusus, kami menganalisis tujuh model pembelajaran mendalam, termasuk, jaringan saraf dalam, jaringan saraf berulang, jaringan saraf convolutional, mesin Boltzmann terbatas, jaringan kepercayaan mendalam, Boltzmann dalam mesin, dan autoencoder yang dalam. Untuk setiap model pembelajaran mendalam, kami mempelajari kinerja model dalam klasifikasi biner dan klasifikasi multiklas.	Dalam makalah ini, kami melakukan studi banding teknik pembelajaran yang mendalam untuk deteksi intrusition, yaitu, model diskriminatif yang mendalam dan gen- model cratif tanpa pengawasan. Secara khusus, kami ana- lisis tujuh pendekatan pembelajaran yang mendalam, termasuk, jaringan saraf dalam, jaringan saraf berulang, jaringan saraf convolutional, Boltzmann terbatas mesin, jaringan kepercayaan yang dalam, mesin Boltzmann yang dalam, dan autoencoder yang dalam.	-	Penelitian yang dilakukan pada ini adalah keamanan cyber, dan pada penelitan yang saya lakukan mendeteksi serangan brute force

Tabel 2.1. Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
7	PERFORMANCE ANALYSIS COMPARISON ON VARIOUS CYBER-ATTACK DATASET BY RELATING A DEEP BELIEF NETWORK MODEL ON AN INTRUSION DETECTION SYSTEM (IDS)	S.Priya 1, Dr.K.Pradeep Mohan Kumar 2, Journal, 2021	Tujuan dari peneliti adalah teknik pembelajaran mendalam sedang diterapkan pada menekankan aplikasi keamanan jaringan. Paling banyak diamati dalam sistem keamanan jaringan adalah penyusup, yang dapat virus, serangan Dos, dan Penetrasi di antara jaringan membuat perbedaan dalam aktivitas jaringan metode dinamis dapat diikuti untuk mendeteksi dan mencegah serangan oleh penyusup. Dalam istilah, deteksi intrusi system(IDS) memiliki begitu banyak kumpulan data statis yang dianalisis untuk alinyemen lalu lintas. Dalam aspek itu untuk lebih akurasi dan analisis teknik pembelajaran mendalam.	Dalam makalah ini, analisis kinerja dataset IDS memiliki: telah direpresentasikan bersama dengan analisis fiturnya dengan membandingkan berbagai kumpulan data yang diterapkan melalui jaringan kepercayaan yang mendalam bersama dengan pengklasifikasi Sigmoid. Untuk menyoroti fokus kami pada Dataset UNSW-NB15 serangan yang paling dikenal seperti negara adalah perubahan yang tidak normal.	dapat diklasifikasikan berdasarkan model DBF dan diterapkan sigmoid pengklasifikasi yang membantu mengidentifikasi perilaku yang tepat dari setiap fitur pada variabelnya menemukan pemodelan yang sesuai	Penelitian ini hanya mendeteksi serangan brute force dengan menggunakan deep belief network.

Tabel 2.1. Lanjutan

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
8	BETH Dataset: Real Cybersecurity Data for Anomaly Detection Research	Kate Highnam * 1 Kai Arulkumaran * 2 1 Zachary Hanif * 3 Nicholas R. Jennings 1, Paper, 2021	Kami menyajikan dataset keamanan siber BETH untuk deteksi anomali dan analisis di luar distribusi. Dengan "anomali" nyata yang dikumpulkan menggunakan sistem pelacakan baru, dataset kami berisi lebih dari delapan juta titik data yang melacak 23 host. Setiap host telah menangkap aktivitas jinak dan, paling banyak, satu serangan, memungkinkan analisis perilaku yang lebih bersih.	kami menyajikan dataset keamanan siber BETH kami untuk deteksi anomali dan analisis OoD. Data bersumber dari sistem pelacakan honeypot baru kami yang merekam peristiwa proses tingkat kernel dan lalu lintas jaringan DNS. Ini berisi serangan dunia nyata di hadapan OS modern yang jinak dan lalu lintas penyedia cloud, tanpa kompleksitas tambahan dari aktivitas pengguna buatan yang bising.	kami berencana untuk mengumpulkan dan mempublikasikan lebih banyak serangan untuk dataset pengujian alternatif. Ini juga akan memungkinkan penyelidikan dalam membandingkan serangan atau mungkin pengujian dalam pengaturan pembelajaran berkelanjutan.	Pada penelitian ini saya menggunakan dataset dari paper tersebut untuk melakukan percobaan, dengan menggunakan metode deep belief network.

2.3. Landasan Teori

2.3.1 *Deep Belief Network* (DBN)

Deep Belief Network (DBN) adalah teknik generatif. DBN terdiri dari RBM bertumpuk yang melakukan pelatihan berlapis-lapis untuk mencapai eksekusi yang solid dalam domain yang tidak diawasi.(Manimurugan et al. 2020) Dalam DBN, pelatihan dicapai lapis demi lapis, dan masing-masing dilakukan sebagai RBM yang dilatih di atas lapisan yang dilatih sebelumnya (DBN adalah sekelompok lapisan RBM yang digunakan untuk tahap pra-pelatihan dan sebagai tambahan diubah menjadi jaringan umpan-maju untuk penyetelan bobot dengan pendekatan yang berbeda. Penggunaan RBM yang signifikan kemungkinan besar karena ada kelangkaan data berlabel, dan RBM serta enkoder otomatis dapat dilatih sebelumnya pada data tak berlabel dan disetel dengan baik pada sejumlah kecil label berlabel. data. Algoritma pelatihan serakah *layer-wise* digunakan untuk melatih DBN satu lapisan pada satu waktu. Metode lapisan bijaksana serakah digunakan karena mengoptimalkan setiap lapisan pada satu waktu dengan rakus. Setelah pelatihan tanpa pengawasan, biasanya ada tahap *finetune*, ketika algoritma pelatihan yang diawasi bersama diterapkan ke semua lapisan. Ini menggabungkan dua ide: 1) bahwa pilihan parameter awal jaringan saraf dalam dapat memiliki efek pengaturan yang signifikan; 2) bahwa pembelajaran tentang distribusi input dapat membantu pembelajaran tentang pemetaan dari input ke output. Pada tahap pra-pelatihan, fitur yang mendasarinya dilatih dengan metode serakah tanpa pengawasan, sedangkan lapisan softmax diimplementasikan dalam tahap

penyetelan halus ke lapisan atas untuk meningkatkan fitur dari sampel berlabel. Gambar.3 mewakili arsitektur DBN.

Untuk merepresentasikan kompleksitas secara visual, kami menstandarisasi devisiasi seperti pada persamaan.1:

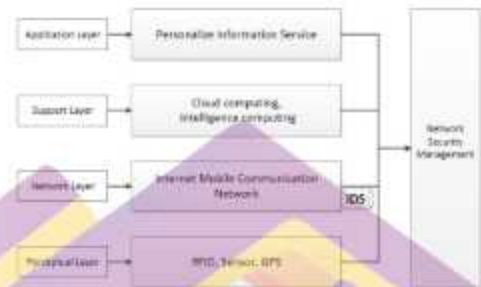
$$\sigma^* = \frac{\sigma - \sigma_{min}}{\sigma_{max} - \sigma_{min}}$$

Dalam RBM, v menunjukkan setiap unit yang terlihat dan h menunjukkan setiap unit yang tersembunyi. Untuk memutuskan sistem, mereka berusaha memperoleh tiga parameter model: = {W, A, B}. Ini adalah matriks bobot W , bias elemen lapisan tersembunyi B, dan bias elemen lapisan terlihat A, masing-masing. Asumsikan sebuah RBM memiliki m sel tersembunyi dan n sel terlihat, vi menunjukkan unit terlihat ke-i, hi unit tersembunyi ke-j, dan struktur parameter ditunjukkan seperti pada persamaan.2:

$$W = \{w_{i,j} \in R^{n \times m}\}$$

2.3.2 Deteksi Intrusi

Deteksi intrusi diterima sebagai sistem keamanan penting yang dirancang untuk mengelola serangan pada jaringan dan mengenali tindakan ganas dalam lalu lintas jaringan komputer.(Manimurugan et al. 2020) Ini mengasumsikan peran penting dalam keamanan data secara keseluruhan dan mendukung dalam menemukan, memutuskan, dan mendeteksi penggunaan, duplikasi, modifikasi, dan pembongkaran kerangka data dan data yang tidak disetujui.



GAMBAR 2.1. Arsitektur IoT.

Kerangka kerja keamanan yang mengamankan jaringan dan sistem dasar dari akses, kerusakan, penghancuran, dan perubahan yang tidak terbukti. Kedua kerangka kerja ini mungkin terdiri dari berbagai model keamanan terkoordinasi; misalnya, *firewall*, antivirus, dan *Intrusion Detection Systems* (IDS) yang memungkinkan jaringan atau sistem untuk diamati dan raise peringatan ketika tindakan ganas terjadi.

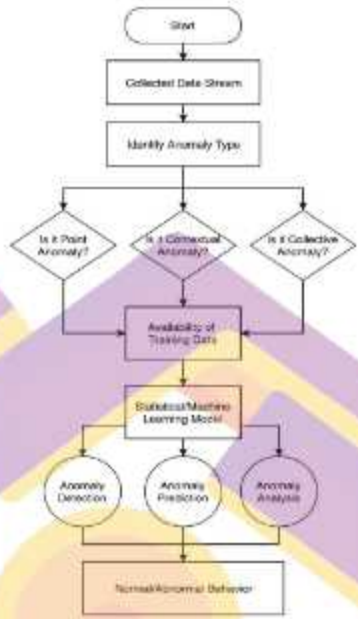
Secara komprehensif, IDS diklasifikasikan menjadi tiga teknik: deteksi penyalahgunaan, deteksi anomali, dan hibrida. Metode identifikasi penyalahgunaan menggunakan tanda tangan yang telah ditentukan dari tindakan ganas untuk mendeteksi intrusi. Oleh karena itu, mereka digunakan untuk mengidentifikasi serangan *known*. Metode deteksi anomali mencirikan pola khas dan mendeteksi tindakan berbahaya tergantung pada perbedaannya dari pola biasa. Dengan cara ini, teknik identifikasi berbasis anomali memiliki kemampuan untuk mengidentifikasi serangan *zero-day*. Metode *hybrid* mengeksploitasi teknik identifikasi anomali dan anomali. Dengan mengurangi positif palsu dari serangan yang tidak diketahui,

metodologi hibrida menargetkan memperluas tingkat identifikasi intrusi yang diketahui.

Deteksi penyusup adalah salah satu advance yang signifikan dalam memastikan keamanan jaringan IoT. Oleh karena itu, deteksi intrusi adalah salah satu dari banyak sistem untuk menangani gangguan keamanan yang dapat diidentifikasi di salah satu dari empat lapisan arsitektur IoT. Lapisan Jaringan tidak hanya berpendapat sebagai dukungan untuk menghubungkan berbagai perangkat IoT; tetapi juga memfasilitasi sistem pertahanan keamanan berbasis jaringan seperti NIDS. Ada banyak teknik IDS; misalnya, teknik yang bergantung pada analisis statistik, analisis kluster, ANN, atau pembelajaran deep. Dalam teknik ini, deteksi intrusi yang bergantung pada pembelajaran mendalam berkinerja lebih baik daripada berbagai teknik lainnya, karena pembelajaran mendalam memiliki kapasitas tinggi untuk belajar mandiri, adaptasi diri, generalisasi, dan identifikasi aktivitas serangan yang tidak diketahui.

2.3.3 Deteksi Anomali

Dunia saat ini memiliki IoT yang luas yang menghasilkan sejumlah besar informasi, dan anomali merupakan bagian penting dari setiap kerangka kerja. Anomali ini bisa menjadi indikasi saluran sumber daya dalam kerangka industri,



GAMBAR 2.2. Bagan alur deteksi anomali.

Keadaan penting di *platform* aeronautika untuk mengabaikan masalah yang tidak terduga, atau mengidentifikasi kinerja instrumen medis yang tidak biasa, dan sebagainya. Oleh karena itu, memiliki opsi untuk mengidentifikasi anomali dapat sangat mempengaruhi kinerja total dari setiap model yang dipantau. Kesulitan utama dalam memahami anomali adalah menggambarkan batas-batas yang tepat antara aktivitas abnormal / normal, karena aksesibilitas observation abnormal untuk melatih model biasanya tidak cukup. Dalam situasi praktis, pola perilaku abnormal telah minimal kontras dengan perilaku normal. Gambar.2 menunjukkan diagram alur deteksi anomali.

Dalam kerangka identifikasi anomali (seperti yang ditunjukkan pada Gambar 2), proses awal adalah untuk memahami kondisi aliran data yang

dikumpulkan, yang cenderung biner, diskrit, atau kontinu, serta kerangka kerja hubungan. Kerangka kerja hubungan ini menunjukkan apakah itu informasi deret waktu, informasi spasial, atau informasi grafis. Mengidentifikasi jenis hubungan mendukung pemilihan metode yang tepat untuk mendeteksi anomali, pemeriksaan, atau harapan. Langkah selanjutnya adalah menemukan jenis anomali dari himpunan yang telah ditentukan (misalnya: anomali titik, anomali kolektif, atau anomali kontekstual).

Proses selanjutnya adalah memahami keberadaan informasi pelatihan untuk merancang kerangka identifikasi anomali. Berdasarkan kehadiran informasi dan penjelasannya, kami mungkin mewakilinya sebagai diawasi, semi-diawasi atau tidak diawasi. Da ta itu membantu pengembang untuk memilih strategi identifikasi anomali yang sesuai. Dalam pelatihan yang diawasi, ketersediaan informasi dengan label kelas dan type dasar pembelajarannya digunakan untuk mengidentifikasi perilaku abnormal kerangka kerja. Dalam pembelajaran tanpa pengawasan, kami memiliki informasi tetapi tidak ada output yang *solid* (misalnya, label kelas). Juga, dalam pembelajaran semi-diawasi kami memiliki model terbatas dengan label kelas sementara informasi yang tersisa tidak diberi label.

2.3.4 JENIS ANOMALI

Bagian penting dari metode identifikasi anomali adalah konsep anomali yang diperlukan. Anomali dapat dicirikan menjadi tiga klasifikasi berikut:

Anomali Titik: Jika model data individu dapat diperlakukan sebagai abnormal mengenai informasi remaining, model melibatkan anomali titik. Ini adalah jenis anomali yang paling dasar dan telah menjadi target sebagian besar

analisis pada identifikasi anomali. Dalam contoh praktis, identifikasi penipuan kartu kredit, biarkan dataset dibandingkan dengan pertukaran kredit seseorang. Untuk menyederhanakan berbagai hal, mari kita terima bahwa informasi ditentukan hanya dengan menggunakan satu fitur: jumlah yang dihabiskan. Transaksi tunai yang lebih tinggi dari tingkat biasa yang akan dihabiskan individu adalah anomali poin. (Manimurugan et al. 2020)

Anomali konteksual: Jika model data tidak normal dalam konteks tertentu (tetapi tidak dalam konteks lain), itu berisi anomali kontekstual (juga dinyatakan sebagai anomali bersyarat). Konsep konteks disebabkan oleh struktur dalam dataset dan harus dihalangi sebagai segmen definisi masalah. Setiap model data dicirikan menggunakan dua pengaturan fitur yang menyertainya. Atribut kontekstual digunakan untuk menentukan konteks (atau lingkungan) untuk model itu; misalnya, dalam data spasial, garis lintang dan lintang suatu daerah adalah atribut kontekstual. Dalam informasi deret waktu, waktu adalah atribut kontekstual yang menentukan kondisi model pada total pesanan.

Atribut Perilaku: Ini mencirikan atribut nonkontekstual suatu model; misalnya, dalam dataset spasial yang mendefinisikan curah hujan rata-rata seluruh dunia, ukuran curah hujan di daerah mana pun adalah atribut perilaku. Perilaku anomali diselesaikan dengan memanfaatkan kualitas untuk atribut perilaku dalam konteks tertentu. Model data dapat mewakili anomali kontekstual dalam konteks yang disediakan. Namun, model data yang setara (sejauh atribut perilaku pergi) akan dipandang sebagai biasa dalam konteks yang berbeda. Properti ini adalah kunci dalam *reognizing* atribut perilaku dan kontekstual untuk metode pengenalan

anomali kontekstual. Keputusan untuk menerapkan metode identifikasi anomali kontekstual dibuat dengan mengenali pentingnya anomali kontekstual di area *application* objektif. (Manimurugan et al. 2020)

Anomali Kolektif: Jika akumulasi model data yang berkaitan adalah anomali mengenai seluruh dataset, itu dikenal sebagai anomali kolektif yang mengandung. Model data individu dalam anomali ini mungkin tidak anomali dengan sendirinya, tetapi peristiwanya bersama sebagai akumulasi adalah anomali. Anomali kolektif diselidiki untuk pengaturan, grafik, dan informasi spasial. Perlu dicatat bahwa sementara anomali titik dapat terjadi dalam dataset apa pun, anomali kolektif hanya dapat terjadi dalam dataset di mana model data terhubung. Perbedaannya adalah bahwa peristiwa anomali kontekstual bergantung pada aksesibilitas atribut konteks dalam informasi. Anomali kolektif atau titik juga bisa menjadi anomali kontekstual setiap kali dipertimbangkan dalam kaitannya dengan konteks. Dengan cara ini, masalah identifikasi anomali titik atau kolektif dapat menjadi masalah identifikasi anomali kontekstual melalui konsolidasi informasi konteks.

2.4 CSE-CIC-IDS2018 (Communications Security Establishment (CSE) & the Canadian Institute for Cybersecurity (CIC))

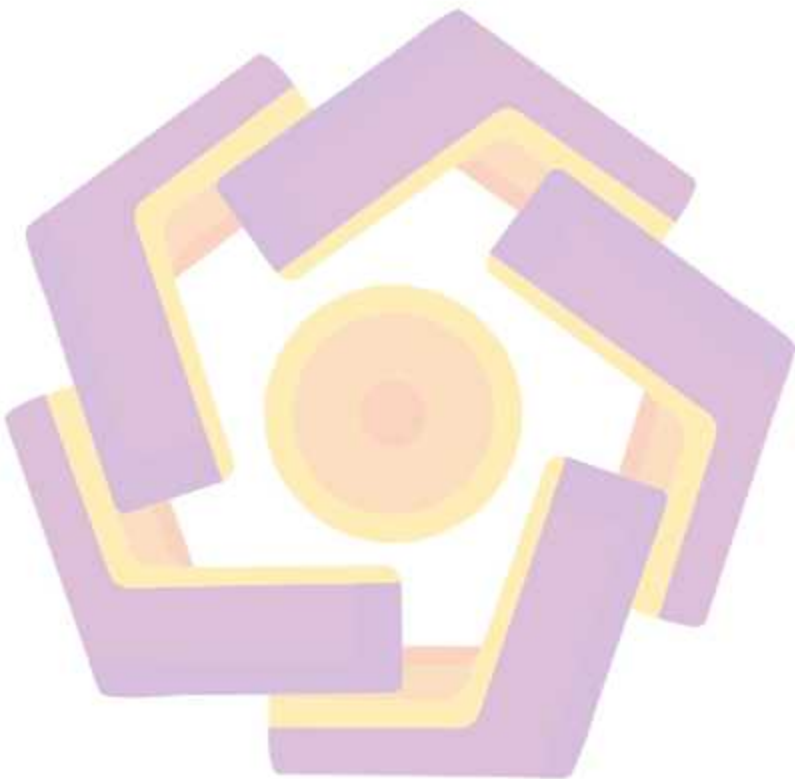
Deteksi anomali telah menjadi fokus utama banyak peneliti karena potensinya dalam mendeteksi serangan baru. Namun, pengadopsiannya ke aplikasi dunia nyata telah terhambat karena kompleksitas sistem karena sistem ini memerlukan sejumlah besar pengujian, evaluasi, dan penyetelan sebelum

penerapan. Menjalankan sistem ini melalui jejak jaringan berlabel nyata dengan serangkaian intrusi dan perilaku abnormal yang komprehensif dan ekstensif adalah metodologi yang paling idealis untuk pengujian dan evaluasi.

Ini sendiri merupakan tantangan yang signifikan, karena ketersediaan kumpulan data sangat jarang, karena dari satu sisi, banyak kumpulan data semacam itu bersifat internal dan tidak dapat dibagikan karena masalah privasi, dan di sisi lain yang lain sangat dianonimkan dan tidak mencerminkan arus. tren, atau tidak memiliki karakteristik statistik tertentu, sehingga kumpulan data yang sempurna belum ada. Dengan demikian, peneliti harus menggunakan kumpulan data yang seringkali kurang optimal. Ketika perilaku dan pola jaringan berubah dan intrusi berkembang, sangat penting untuk beralih dari kumpulan data statis dan satu kali ke kumpulan data yang dihasilkan secara lebih dinamis, yang tidak hanya mencerminkan komposisi lalu lintas dan intrusi pada waktu itu, tetapi juga dapat dimodifikasi, dapat diperluas, dan dapat direproduksi. Untuk mengatasi kekurangan ini, pendekatan sistematis telah dirancang untuk menghasilkan kumpulan data untuk menganalisis, menguji, dan mengevaluasi sistem deteksi intrusi, dengan fokus pada detektor anomali berbasis jaringan.

Serangan brute force: Serangan brute force: Serangan brute force sangat umum terhadap jaringan karena mereka cenderung membobol akun dengan kombinasi nama pengguna dan kata sandi yang lemah. Skenario terakhir telah

dirancang dengan tujuan memperoleh akun SSH dan MySQL dengan menjalankan serangan brute force kamus terhadap server utama.



BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

i. Jenis Penelitian

Jenis Penelitiannya adalah simulasi. Penelitian ini merupakan penelitian lanjutan dari wanjau (2021), yang mana penelitian itu memiliki data di dalam jurnalnya dan dibandingkan dengan hasil deep learning yaitu Deep Belief Network yang dikerjakan di penelitian ini. Disini bermaksud untuk mengetahui deep model mana yang bagus dalam menganalisa serangan dari brute force.

ii. Sifat Penelitian

Dilihat dari sifat penelitian ini adalah penelitian kuantitatif, artinya hasil dari penerapan metode-metode untuk mengklasifikasi data berupa angka untuk mengetahui metode yang menghasilkan tingkat akurasi, presisi, recall, dan f1 score yang baik dan menganalisis performa suatu model dalam mengklasifikasikan serangan brute force.

iii. Pendekatan Penelitian Kuantitatif

Penelitian ini menggunakan pendekatan kuantitatif yaitu hasil dari penerapan metode-metode untuk mengklasifikasikan data berupa angka dan diagram untuk menunjukkan tingkat akurasi dan performa daripenerapan metode-metode tersebut.

3.2. Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari dataset IDS 2018 intrusion CSVs (CSV-CIC-IDS2018) yang bersumber dari Kaggle. data ini digunakan karena, ini adalah salah satu dari data runjukan dari jurnal internasional sebelumnya yang meneliti dari ISCX IDS 2012. Dan dataset yang digunakan adalah dataset IDS 2018 intrusion CSVs (CSV-CIC-IDS2018) yang didalam dataset tersebut memiliki beberapa serangan dari brute force.

3.3. Metode Analisis Data

Metode analisis yang digunakan adalah *Deep Belief Network* (DBN), yang terdiri dari Restricted Boltzman Machine yang melakukan pelatihan layer-wise untuk mencapai eksekusi yang solid dalam domain tanpa pengawasan. Penggunaan RBM yang signifikan kemungkinan akan terjadi, ada kelangkaan data berlabel, dan RBM dan encoder otomatis dapat dilatih sebelumnya pada data yang tidak berlabel dan disesuaikan pada sejumlah kecil data berlabel. Asumsikan sebuah RBM memiliki m sel tersembunyi dan n sel terlihat, vi menunjukkan unit ke-i yang terlihat, hai unit tersembunyi ke-j, dan struktur parameter ditunjukkan seperti pada persamaan.2:

$$W = \left\{ w_{i,j} \in R^{n \times m} \right\} \tag{2}$$

Di mana wi,j menunjukkan bobot di antara sel terlihat ke-i dan sel tersembunyi ke-j dari persamaan 3.

$$A = \{a_i \epsilon R^m\}$$

(3)

Di mana, ai mewakili ambang bias dari sel yang terlihat ke-i dari persamaan 4;

$$A = \{b_j \epsilon R^n\}$$

(4)

Di mana, bj menunjukkan ambang batas bias sel yang terlihat ke-j. Untuk urutan (v, h) melalui kondisi sekarang, dengan asumsi lapisan tersembunyi dan terlihat mengikuti distribusi Bernoulli, persamaan energi RBM direpresentasikan

$$E(v, h \mid \theta) = -\sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{m} b_j h_j - \sum_{i=1}^{n} \sum_{j=1}^{m} v_i W_{ij} h_j \quad (5)$$

seperti pada persamaan 5:

Dimana, Θ = {Wij, ai, bj} adalah parameter model RBM, dan fungsi energi menunjukkan nilai energi antara estimasi setiap simpul yang terlihat dan setiap lapisan tersembunyi simpul. Karena regularisasi dan eksponensial energi fungsi, persamaan distribusi kemungkinan bersama bisa menjadi diperoleh di mana node mengatur lapisan yang terlihat dan *node set* lapisan tersembunyi berada dalam kondisi tertentu secara terpisah (v,h) seperti pada persamaan 6:

$$P(v,h\,|\,\theta) = \frac{e^{-E(v,h\,|\,\theta)}}{Z(\theta)} \tag{6}$$

$$Z(\theta) = \sum_{v,h} e^{-E(v,h\,|\,\theta)} \tag{7}$$

Di mana, dalam persamaan 7, Z(θ) adalah faktor standar atau fungsi distribusi yang menunjukkan eksponen energi total dari setiap kondisi yang tersedia dari himpunan node tersembunyi dan lapisan yang terlihat.

Penentuan fungsi probabilitas sering kali digunakan untuk mendapatkan parameter Setelah mempresentasikan bersama distribusi kemungkinan P(v,h|θ), distribusi marginal P(v|θ) dari kumpulan node dari lapisan yang terlihat dapat diperoleh melalui penjumlahan dari keseluruhan kondisi yang tersembunyi node lapisan diatur dalam persamaan 8:

$$P(v\,|\,\theta) = \frac{1}{Z(\theta)} \sum_{h} e^{-E(v,h\,|\,\theta)} \tag{8}$$

Distribusi marjinal menunjukkan kemungkinan dengan di mana susunan node di lapisan yang terlihat adalah dalam distribusi tingkat tertentu. Karena kecuali- koneksi lapisan-lapisan nasional dan tanpa koneksi antar-lapisan bentuk sistem RBM, ia memiliki signifikansi yang menyertainya kondisi: Setelah mempresentasikan kondisi sel yang terlihat, kondisi berlakunya setiap sel lapisan

tersembunyi adalah otonom terbatas. Di sini, kemungkinan inisiasi dari elemen tersembunyi ke-jth adalah seperti yang ditunjukkan pada persamaan 9:

$$P(h_j = 1 \mid v) = \sigma(b_j + \sum_i v_i W_{ij}) \tag{9}$$

Dengan demikian, setelah kondisi elemen tersembunyi adalah ditentukan, kemungkinan inisiasi dari elemen yang terlihat tambahan independen bersyarat seperti yang diwakili dalam persamaan 10:

$$P(v_i = 1 \mid h) = \sigma(a_i + \sum_j W_{ij} h_j) \tag{10}$$

Dimana, ô(x) adalah fungsi sigmoid. Untuk memutuskan model RBM, penting untuk menyortir keluarkan tiga parameter model: = {Wij, ai, bj}.Susunan parameter menggunakan probability berfungsi untuk mengambil parameter bawahan. Dari persamaan 8, energi E berbanding terbalik dengan peluang P, dan E terbatas melalui perluasan P. Strategi reguler untuk memperluas masalah fungsional Kemampuan adalah teknik menaikkan kemiringan yang berhubungan dengan perubahan parameter seperti yang ditunjukkan oleh yang menyertai persamaan 11:

$$\theta = \theta + \mu \frac{\partial \ln P(v)}{\partial \theta} \tag{11}$$

Proses berulang ini memperluas probabilitas P dan mengurangi energi E. Aliran algoritma dapat diuraikan sebagai:

Langkah 1: Memulai populasi dan menghasilkan beragam lapisan tersembunyi dan total neuron di setiap lapisan secara acak;

Langkah 2: Hitung tingkat kebugaran sesuai Eq. 1, dipilih oleh teknik roulette, dan pertahankan individu ideal di masa sekarang; *interval crossover*; variasi;

Langkah 3: "Elite" memegang, memegang individu dengan nilai kebugaran terbaik dalam pengembangan proses;

Langkah 4: Temukan apakah jumlah iterasi tertinggi telah tercapai. Setelah tercapai, struktur jaringan yang dihasilkan adalah *held*, atau ulangi Step2-Step3 sekali lagi;

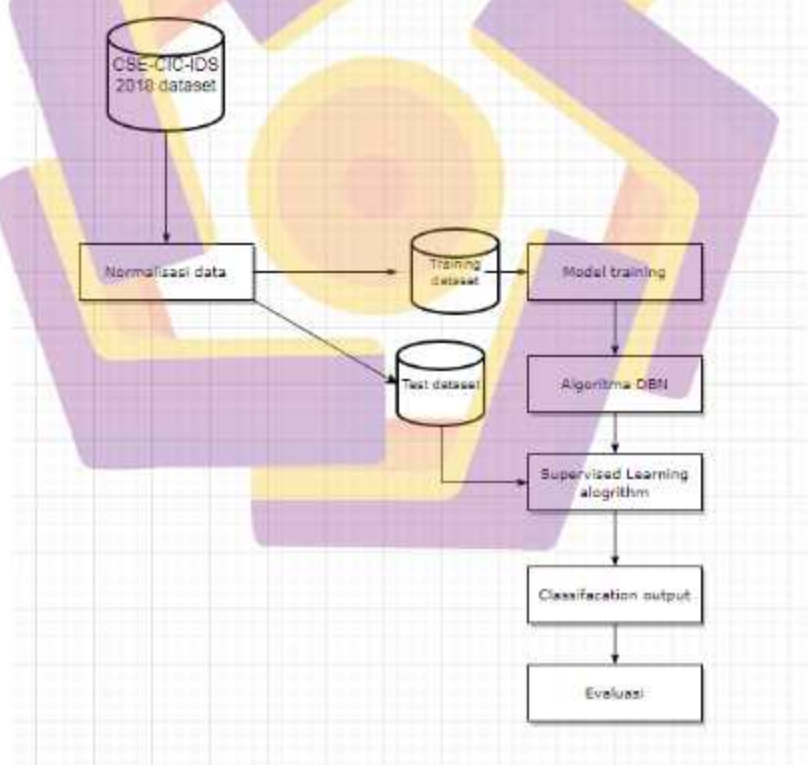
Langkah 5: Memanfaatkan struktur jaringan yang optimal untuk DBN dan melatih sistem IDS;

Langkah 6: Mengklasifikasikan set pengujian melalui model DBN terlatih, dan terakhir mengoordinasikan hasil klasifikasi dengan data *classification* dari set pengujian untuk memvalidasi akurasi klasifikasi.

3.4. Alur Penelitian

Alur pertama didalam penelitian adalah mengumpulkan data, lalu ketika sudah terkumpul data-datanya, peneliti mengindentifikasi data-data tersebut

menjadi 3 pada data anomali tersebut. Yaitu data titik anomali, data anomali konstektual, dan data anomali kolektif. Ketika ketiga data anomali tersebut selesai di indefikasi lalu semua data tersebut disatukan menjadi CSV (*Command Separated Value*). Setelah Data CSV tersedia kita inputkan ke dalam machine learning pada DBN untuk mengetahui hasilnya. Setelah mendapatkan hasilnya, dapat dibandingkan dengan hasil data dari peneliti sebelumnya dan memilah dari beberapa algoritma tersebut agar mengetahui hasil yang cukup baik untuk mendeteksinya. Seperti pada gambar 3 *flowchart* dibawah ini.



Gambar 3.1. Flowchart Program DBN

Dari gambar diatas, adalah flowchart proses dari kinerja DBN, pada Dataset CSE-CIC_IDS 2018 akan dinormalisasikan datanya, ketika sudah di normalisasikan akan dibagi 2 proses yaitu training dataset dan test dataset. Pada training dataset dibagikan 2 yaitu model training dan algoritma DBN, di model training ini akan memilah hasil dataset yang ada dan meanalisis tiap label didalamnya dan algoritma DBN ini adalah algoritma yang digunakan peneliti didalam penelitian ini yang akan di uji coba menggunakan dataset yang di uji.

Test dataset ini terbagi lagi menjadi 3 yaitu Supervised Learning allgorithm, classification output, dan evaluasi. Pada supervised learning algorithm disini peneliti menggunakan pemograman di google colab untuk menjalankannya dengan menggunakan algoritma DBN dan CNN deep learning untuk menjalankannya. Untuk classifacation output pembagian class yang data-data di dalamnya dari output portnya maupun total dari semua data-data didalamnya. Evaluasi adalah hasil dari program yang kita jalanin dan apakah ada kesalahan didalamnya ketika menjalankannya.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Dataset

Dalam penelitian ini data yang diteliti adalah deep model yaitu deep belief network dataset yang digunakan adalah IDS 2018 intrusion CSVs (CSV-CIC-IDS2018) di dataset tersebut memiliki berbagai serangan didalamnya, penelitian ini hanya berfokus ke serangan brute force saja. Komposisi didalam dataset (CSVs-CIC-IDS2018) adalah iris dataset yang berarti dimachine learning yang digunakan untuk klasifikasi data. Yang didalamnya memiliki beberapa sampel didalamnya.

Didalam dataset tersebut memiliki data 1048576 data terdiri atas benign (jinak) sekitar 665355, setelah itu pada serangan FTP Brute Force 193354 dan SSH-Brute Force memiliki sekitar 187589. Didalam data tersebut akan diambil serangan *FTP brute force* dan *SSH brute force* kemudian setelah data telah terkumpul akan di prepocessing ke dalam google colaboratory.

Tabel 4.1 Penjelasan komposisi atribut dataset

Dst Port	Destination port of Connection
Protocol	Protocol used during Connecion
Timestamp	Time that connection occured
Flow Duration	Duration that connection occured
Tot fwd Duration	Total number of forward packets.

Tot bwd Pkts	Total number of backward packets.
...	...
Totlen fwd Pkts	Total length forward packet
Totlen bwd Pkts	Total legth backward packet
Fwd len max	Maximun legth packet
Fwd len min	Minimun legth packet

Komposisi dataset memiliki 80 data atribut. Terdiri atas Index (['Dst Port', 'Protocol', 'Timestamp', 'Flow Duration', 'Tot Fwd Pkts', 'Tot Bwd Pkts', 'TotLen Fwd Pkts', 'TotLen Bwd Pkts', 'Fwd Pkt Len Max', 'Fwd Pkt Len Min', 'Fwd Pkt Len Mean', 'Fwd Pkt Len Std', 'Bwd Pkt Len Max', 'Bwd Pkt Len Min', 'Bwd Pkt Len Mean', 'Bwd Pkt Len Std', 'Flow Byts/s', 'Flow Pkts/s', 'Flow IAT Mean','Flow IAT Std', 'Flow IAT Max', 'Flow IAT Min', 'Fwd IAT Tot','Fwd IAT Mean', 'Fwd IAT Std', 'Fwd IAT Max', 'Fwd IAT Min', 'Bwd IAT Tot', 'Bwd IAT Mean', 'Bwd IAT Std', 'Bwd IAT Max','Bwd IAT Min', 'Fwd PSH Flags', 'Bwd PSH Flags', 'Fwd URG Flags','Bwd URG Flags', 'Fwd Header Len', 'Bwd Header Len', 'Fwd Pkts/s','Bwd Pkts/s', 'Pkt Len Min', 'Pkt Len Max', 'Pkt Len Mean','Pkt Len Std', 'Pkt Len Var', 'FIN Flag Cnt', 'SYN Flag Cnt','RST Flag Cnt', 'PSH Flag Cnt', 'ACK Flag Cnt', 'URG Flag Cnt','CWE Flag Count', 'ECE Flag Cnt', 'Down/Up Ratio', 'Pkt Size Avg','Fwd Seg Size Avg', 'Bwd Seg Size Avg', 'Fwd Byts/b Avg','Fwd Pkts/b Avg', 'Fwd Blk Rate Avg', 'Bwd Byts/b Avg','Bwd Pkts/b Avg', 'Bwd Blk Rate Avg', 'Subflow Fwd Pkts', 'Subflow Fwd Byts', 'Subflow Bwd Pkts', 'Subflow Bwd Byts','Init Fwd Win Byts', 'Init Bwd Win Byts', 'Fwd Act Data Pkts', 'Fwd Seg Size

Min', 'Active Mean', 'Active Std', 'Active Max','Active Min', 'Idle Mean', 'Idle Std', 'Idle Max', 'Idle Min', 'Label'],dtype='object'). Peneliti memasukan beberapa komposisi agar pembaca dapat mengetahui beberapa pengertian komposisi didalamnya.

Jenis kelas yang diperdiksi didalam ini adalah Deep Belief network yang dari sub-bagian DNN (Deep Neural Network) Inti perhitungan pada algoritma berbasis jaringan adalah untuk mencari bobot terbaik dari contoh / sampel data yang sudah ada. Karena hasil pada contoh data sudah diketahui, maka nilai bobot akan dihitung berdasarkan nilai hasil yang sudah tersedia, sampai ditemukan nilai bobot terbaik yang paling banyak cocok apabila dihitungkan kembali pada data awal. Kemudian nilai bobot tersebut dapat digunakan untuk menghitung data lain yang tidak diketahui hasilnya. Pada kasus ini, metode yang digunakan untuk mencari bobot terbaik adalah menggunakan metode PSO (Particle Swarm Optimization).

4.2 Pre-processing

Normalisasi data Untuk mengatasi kelemahan ini, pendekatan sistematis telah dirancang untuk menghasilkan kumpulan data untuk menganalisis, menguji, dan mengevaluasi sistem deteksi intrusi, dengan fokus pada detektor anomali berbasis jaringan. Tujuan utama dari proyek ini adalah untuk mengembangkan pendekatan sistematis untuk menghasilkan kumpulan data tolok ukur yang beragam dan komprehensif untuk deteksi intrusi berdasarkan pembuatan profil pengguna yang berisi representasi abstrak dari kejadian dan perilaku yang terlihat di jaringan. Profil tersebut akan digabungkan untuk menghasilkan kumpulan kumpulan data

yang beragam, masing-masing dengan sekumpulan fitur unik, yang mencakup sebagian dari domain evaluasi.

4.2.1 Pembagian testing dan training

Deteksi anomali telah menjadi fokus utama banyak peneliti karena potensinya dalam mendeteksi serangan baru. Namun, pengadopsiannya ke aplikasi dunia nyata telah terhambat karena kompleksitas sistem karena sistem ini memerlukan sejumlah besar pengujian, evaluasi, dan penyetelan sebelum penerapan. Menjalankan sistem ini melalui jejak jaringan berlabel nyata dengan serangkaian intrusi dan perilaku abnormal yang komprehensif dan ekstensif adalah metodologi yang paling idealis untuk pengujian dan evaluasi.

Ini sendiri merupakan tantangan yang signifikan, karena ketersediaan kumpulan data sangat jarang, karena dari satu sisi, banyak kumpulan data semacam itu bersifat internal dan tidak dapat dibagikan karena masalah privasi, dan di sisi lain yang lain sangat dianonimkan dan tidak mencerminkan arus. tren, atau tidak memiliki karakteristik statistik tertentu, sehingga kumpulan data yang sempurna belum ada. Dengan demikian, peneliti harus menggunakan kumpulan data yang seringkali kurang optimal. Ketika perilaku dan pola jaringan berubah dan intrusi berkembang, sangat penting untuk beralih dari kumpulan data statis dan satu kali ke kumpulan data yang dihasilkan secara lebih dinamis, yang tidak hanya mencerminkan komposisi lalu lintas dan intrusi pada waktu itu, tetapi juga dapat dimodifikasi, dapat diperluas, dan dapat direproduksi.

Untuk mengatasi kekurangan ini, pendekatan sistematis telah dirancang untuk menghasilkan kumpulan data untuk menganalisis, menguji, dan mengevaluasi sistem deteksi intrusi, dengan fokus pada detektor anomali berbasis jaringan. Tujuan utama dari proyek ini adalah untuk mengembangkan pendekatan sistematis untuk menghasilkan kumpulan data tolak ukur yang beragam dan komprehensif untuk deteksi intrusi berdasarkan pembuatan profil pengguna yang berisi representasi abstrak dari kejadian dan perilaku yang terlihat di jaringan. Profil tersebut akan digabungkan untuk menghasilkan kumpulan kumpulan data yang beragam, masing-masing dengan sekumpulan fitur unik, yang mencakup sebagian dari domain evaluasi.

4.2.2 Pre- Processing didalam Google colaboration

Pada proses pertama dalam koding di google colaboration adalah melakukan penginputan file, disini peneliti mengunakan kaggle login untuk bisa menjalankan file yang di kaggle secara langsung. Selanjutnya mengisntal matplotlib serta beberapa libary didalamnya, lalu mengimport file dataset dari kaggle ke dalam google colaboratory, lalu mengekstrak kedalam "tmp" folder didalam google colab. Setelah itu, meload data kedalam memory, lalu menjumlah bentuk didalam dataset tersebut, dengan number of row : 1048575 dan number of column : 80. Lalu pada google colaboration akan menggunakan rumus algoritma dari DBN. Pertama-tama penelitian menginputkan source code packet kedalam google colaboration agar dapat menjalankan program DBN didalam simulasi ini.

Ketika sudah selesai menginstalisasi packetnya, selanjutnya menginstal packet skylearn agar dapat mencari hasil data perhitungan didalamnya. Dan tidak lupa mengekstrak libary agar dapat berjalan. Lalu untuk menjalankan DBN perlu memasukan source code ; classifier = SupervisedDBNClassification untuk mengklasifikasi data didalam dataset yang digunakan. Dengan clasification x train dan y train. Selanjutnya mencari hasil dari dataset tersebut menggunakan DBN yang seperti telah di tentukan di penelitian ini untuk menentukan hasil dari *accuracy, recall, precession, dan f1 Score*. Setelah itu peneliti mencari standart devitiation didalam DBN tersebut dengan mendapatkan hasil 0.24 accuracy with a standard deviation of 0.08.

Menampilkan hasil dari dataset tersebut untuk menunjukan nilai dari akurasi, precesion, recall, dan f1-score. Setelah mendapatkan hasil dari simulasi tersebut maka dapat dibandingkan dengan hasil yang telah didapatkan dengan penelitian sebelumnya.

Tabel 4.2 Network data

	0	1	2	3
Dst port	0	0	0	22
Protocol	0	0	0	6
Timestamp	14/02/2018 08:31:01	14/02/2018 08:33:50	14/02/2018 08:36:39	14/02/2018 08:40:13
Flow Duration	11264 1719	11264 1466	11263 8623	64539 66
Tot fwd pkts	3	3	3	15
Tot bwd pkts	0	0	0	10
Totlen fwd pkts	0	0	0	1239
Totlen bwd pkts	0	0	0	2273
....	...	...	...	...
Fwd pkts len max	0	0	0	744

Tabel 4.3 lanjutan

Fwd pkts len min	0	0	0	0
Idle mean	563208 59.5	563207 33.0	563193 11.5	0.0
Idle std	139.300 036	114.551 299	301.934 596	0.000 000
Idle max	563209 58	563208 14	563195 25	0
Idle min	563207 61	563206 52	563190 98	0
Label	Benign	Benign	Benign	Benign

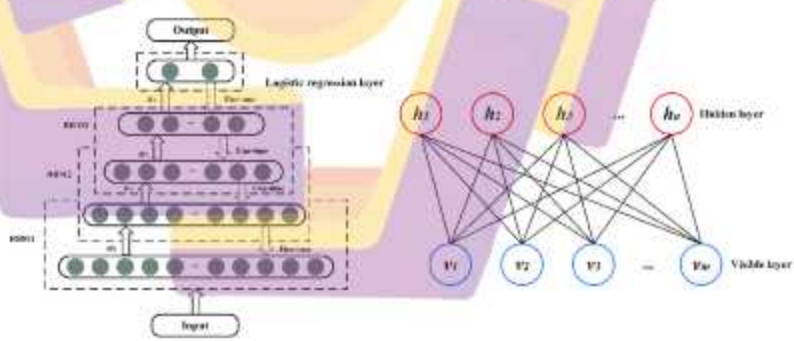
Tabel diatas adalah network data yang ada didalam dataset yang diuji, disini peneliti mengambil hanya mengambil 4 sample data didalamnya dikarenakan sample data yang ada didalam dataset tersebut memiliki sekitar 1.048.575 sample dan dengan column sekitar 80. Dan dataset tersebut dari timestamp 14/02/2018, flow duration yang beragam dari setiap datanya. Total forward paket dari 0-2 memiliki 3 paket dan dinomor 3 memiliki 15 paket didalamnya. Total backward paketnya ada 10 didalam 10. Total legth forward paket 1239, total legth backward 2273, forward paket legth max 744 di no 3. Setelah itu network data column yang

ada di dalam dataset tersebut seperti dst port, protocol, timestamp, flow duration, tot fwd packet, etc. Number column didataset tersebut berisi total 80 data, di bawah data column tersebut ada data info untuk number of columnnya. Selanjutnya mengcheck number value for label didalam datasetnya, benign ; 667626, FTP-Brute Force ; 193360, SSH-Brute Force ; 187589, name ; label, dtype ; int64.

Infrastruktur penyerang mencakup 50 mesin dan organisasi korban memiliki 5 departemen dan mencakup 420 mesin dan 30 server. Kumpulan data mencakup lalu lintas jaringan yang ditangkap dan log sistem dari setiap mesin, bersama dengan 80 fitur yang diekstraksi dari lalu lintas yang ditangkap menggunakan CICFlowMeter-V3.

4.3 Processing

Arsitektur DBN



Gambar 4.1 arsitektur DBN

Dari gambar diatas terlihat ada proses input dan outputnya dari proses input awal akan memasukin tempat layer RBM1, RBM1 adalah awal dari suatu

pengambilan keputusan, sampai ke RBM2 akan diyakinkan lagi bahwa keputusan awal akan dilanjutkan atau tidak. Di RBM3 adalah final result dari keputusan yang dilakukan, setelah RBM3 dilewatin akan memasukin Logistic Reggression untuk memprocessing data yang sedang dicari dan dihubungkan secara matematiks untuk memprediksi nilai dari salah satu faktor tersebut berdasarkan faktor yang lainnya. Dan lingkaran hitam didalam RBM dan Logistic adalah layer yang terbagi menjadi 2 yaitu; visible layer dan hidden layer. Visible layer berfungsi untuk menampung data mentah seperti suara ,teks, atau gambar. Hidden layer berfungsi untuk melatih serangkaian fitur unik berdasarkan output dari jaringan sebelumnya.

Dari gambar tersebut proses untuk mendeteksinya dibutuhkan beberapa Infrastruktur penyerang mencakup 50 mesin dan organisasi korban memiliki 5 departemen dan mencakup 420 mesin dan 30 server. Kumpulan data mencakup lalu lintas jaringan yang ditangkap dan log sistem dari setiap mesin, bersama dengan 80 fitur. Dan ini adalah hasil dari dataset yang diuji yang menggunakan beberapa unit mesin sebagai uji cobanya.

4.4 Hasil Evaluasi

Rumus untuk perhitungan DBN :

Accuracy model dievaluasi dalam hal subset dari kinerja model. Akurasi adalah salah satunya pengukuran untuk menilai model klasifikasi data. Dengan perumusan sebagai berikut :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Presisi menyiratkan tingkat predikatif positif. Ini adalah bagian dari total positif sejati yang dinyatakan model berkorelasi dengan total positif yang dituntutnya. Perumusan presisi sebagai berikut :

$$Precision = \frac{TP}{TP + FP}$$

Recall dikenal sebagai nilai TP, yang mengacu pada total positif dalam sistem menyatakan kontras dengan total yang tepat positif dalam informasi. Perumusan *recall* sebagai berikut :

$$Recall = \frac{TP}{(TP + FN)}$$

F1 *score* juga dapat digunakan untuk memperkirakan kinerja model. Ini adalah rata-rata tertimbang dari recall dan presisi dari modelnya. Perumusan F1 *score* sebagai berikut :

$$F1Score = \frac{2 * TP}{2 * TP + FP + FN}$$

Dalam bab 4 ini peneliti bermaksud untuk membandingkan hasil dari jurnal sebelumnya dan dibandingkan dengan deep learning yaitu Deep Belief network. Dari penelitian sebelumnya mendapatkan suatu hasil didalam penelitiannya.

Peneliti akan membandingkan dari peneliti sebelumnya dari data tabel yang sudah ada dibawah ini, lalu peneliti dapat menyimpulkan hasil dari penelitian ini hasil data mana yang bagus untuk mendeteksi suatu serangan brute force.

Tabel 4.4 Hasil Penelitian Sebelumnya

Kelas	Metrik			
	Akurasi	Precision	Recall	F1-Score
Brute Force	0,943	0,925	0,978	0,918

Tabel diatas adalah hasil dari penelitian sebelumnya. Yang mengambil dari hasil terbaik dari penelitian sebelumnya yaitu CNN yang memiliki nilai berkisar 94,3% Acuracy, 92,5% Precision, 97,8% Recall, dan 91,8% F1-Score.

Tabel 4.5 Hasil Penelitian Menggunakan Fitur Minimum

Kelas	Metrik			
	Akurasi	Precision	Recall	F1-Score
Brute Force	0,852	0,891	0,922	0,894

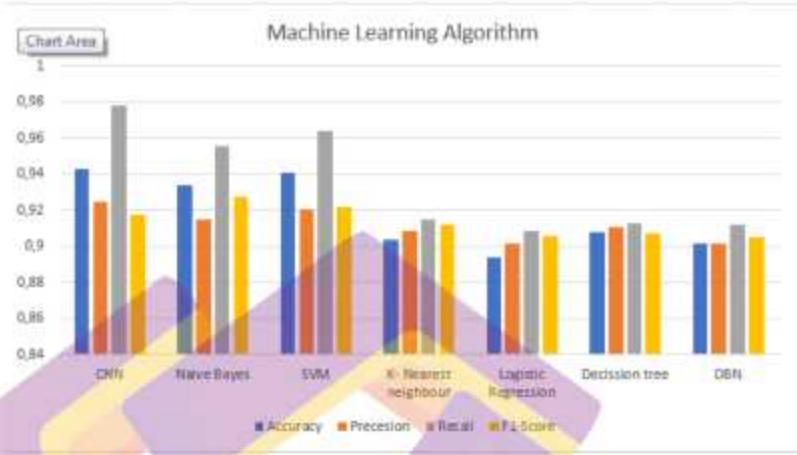
Tabel diatas adalah hasil dari penelitian sebelumnya. Hasilnya menunjukkan bahwa ketika fitur minimal digunakan untuk tugas klasifikasi, kinerja model lebih rendah di semua metrik dibandingkan dengan tugas klasifikasi yang menggunakan semua fitur. Menggunakan fitur minimal untuk

mengklasifikasikan serangan SSH-Brute force, model ini mencapai Acuracy 85,2%, tingkat Presisi 89,2%, tingkat Recall 92,2%, dan skor F1-Score 89,4%.

Tabel 4.6 Hasil Eksperimen DBN di Google Colaboratory

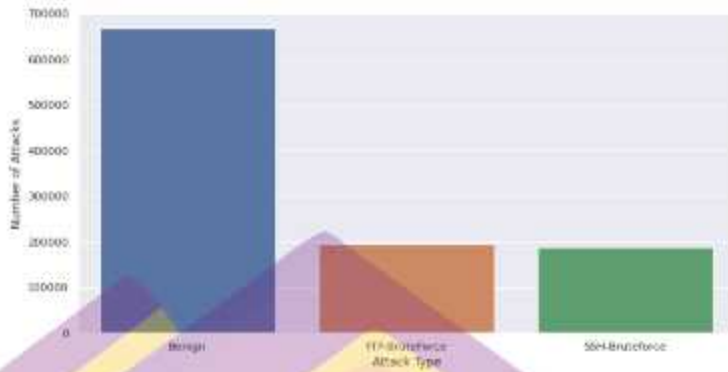
Machine Learning	Accuracy	Recall	Precession	F1-Score
DBN	0.902	0.902	91.674	0.905

Nilai Accuracy pada DBN adalah 0.902, dan untuk recallnya memiliki nilai angka 0.902, nilai dari precession berkisar di angka 91.674, dan nilai angka terakhir yaitu di F1-score berkisaran diangka 0.905. hasil dari nilai tersebut didapatkan di Google Colaborary dan didalam nya tersebut nilainya belum dibuat angka persentase untuk DBN itu sendiri. Jadi untuk mengubah data tersebut untuk menjadi nilai persen peneliti mengambil 4 angka dibelakang koma lalu dikalikan 100% maka akan mendapatkan nilai di Accuracy 90,27%, Recall 90,27%, Precession 91,67%, dan F1-Score 90,51%.



Gambar 4.2. Machine Learning Algorthm

Pada gambar diatas terlihat hasil dari beberapa perbandingan dari hasil Machine learning, hasil terbaik di dapatkan dari hasil CNN, DBN tidak dapat dibandingkan dengan CNN karna memiliki hasil yang sangat berbeda jauh untuk pengukuran hasilnya. Akan tetapi lebih DBN cukup baik dari logistic regression dan hampir sama dengan nilai dari k-nearest neighbour untuk hasilnya. Jadi algoritma DBN ini lebih baik untuk dibandingkan seperti antara logistic regression dan k-Nearest Neighbour untuk melakukan suatu penelitian didalamnya. Dan hasil dari penelitian ini paling terbaik adalah CNN untuk dapat mendeteksi suatu serangan malware ransomware, dan serangan siber lainnya seperti brute Force.



Gambar 4.3. Number Of Attack

Didalam server tersebut memiliki benign (jinak) sekitar 665355, setelah itu pada serangan FTP Brute Force 193354 dan SSH- Brute Force memiliki sekitar 187589. Jadi didalam dataset tersebut memiliki banyak sample data didalamnya akan tetapi untuk serangan brute force ini terbagi 2 hal yaitu FTP dan SSH-Brute Force.

Gambar 4.4. K-fold

Didalam percobaan ini peneliti sudah melakukan pengulangan menggunakan k-cross validation sebanyak 5x pada datasetnya. Akan tetapi hasil data yang didapatkan dari machine learning belum bisa dibilang berhasil dikarenakan nilai yang didapatkan di programnya hanya mendapatkan 0.23 training loss dalam 5x percobaan dan dengan standard deviation 0.04. akan tetapi nilai dari akurasi, recall, precession, dan f1-score yang didapatkan dari data tersebut memiliki hasil yang cukup bagus di Deep Belief network dengan nilai di accuracy 90,27%, recall 90,27%, precession 91,67%, dan F1-score 90,51%. Dari angka tersebut dapat di bandingkan dengan penelitian sebelumnya memiliki hasil nilai yang lebih baik didalamnya.

4.5 Diskusi

Disini peneliti akan menjelaskan apa itu Deep belief network terlebih dahulu. Deep Belief Networks merupakan model deep learning yang memanfaatkan tumpukan/stack Restricted Boltzmann Machines (RBM) atau kadangkala Autoencoders. Autoencoder adalah model neural network yang memiliki input dan output yang sama. Autoencoder mempelajari data input dan berusaha untuk melakukan rekonstruksi terhadap data input tersebut. Peneliti bermaksud menggunakan DBN untuk mengetahui kinerja deep learning yang sudah dianggap lama/kuno untuk mengetahui apakah kinerja dari deep belief network tersebut dapat mengatasi suatu serangan cyber security terutama pada Brute Force yang jadi titik acuan untuk diteliti. Tentu saja, peneliti akan membandingkan dengan kinerja pada peneliti sebelumnya dengan membandingkan hasil dari kinerja mendeteksi suatu serangan khususnya di burte-force.

Didalam dataset number of attack di server tersebut memiliki benign (jinak) sekitar 665355, setelah itu pada serangan FTP Brute Force 193354 dan SSH- Brute Force memiliki sekitar 187589. Data dari penelitian sebelumnya mendapatkan hasil Menggunakan fitur minimal untuk mengklasifikasikan serangan SSH-Brute force, model ini mencapai akurasi 85,2%, tingkat presisi 89,2%, tingkat recall 92,2%, dan skor F1 89,4% yang menggunakan CNN. Dapat disimpulkan dari data yang didapatkan data dari DBN lebih tinggi dengan kisaran dari hasil data penelitian sebelumnya. Akan tetapi di precision data yang di dapatkan dari peneliti sebelumnya memiliki nilai lebih tinggi sekitar 0,53% dari data hasil DBN adalah nilai di accuracy 90,27%, recall 90,27%, precession 91,67%, dan F1-score 90,51%.

Tetapi angka keberhasilan yang bagus itu terletak di CNN yang mengunakan fitur minimal dari peneliti sebelumnya bernilai sekitar 85 % akurasi, 89% presisi, 92% recall, dan 89% f1-score.

BAB V

PENUTUP

5.1. Kesimpulan

Peneliti menyimpulkan pada data dari hasil eksperimen tersebut result DBN itu lebih rendah dari hasil penelitian sebelumnya, akan tetapi dengan mendeteksi suatu serangan menggunakan CNN lebih baik untuk dalam hal mendektsi suatu serangan malware maupun ransomware. Dikarenakan hasil dari eksperimen tersebut menunjukan bahwa tingkat akurasi, recal, precesion, dan f1-score itu lebih baik dari DBN. Dan didalam penelitian sebelumnya menggunakan fitur minimal untuk klasifikasi data.

Parameter yang memperngaruhi kinerja dari deep belief network terletak pada hidden layer dan epoch. Deep Belief network dengan nilai di accuracy 90,27%, recall 90,27%, precession 91,67%, dan F1-score 90,51%. Dapat disimpulkan bahwa hasil dari penelitian ini sudah memenuhi kriteria untuk dapat melakukan pengujian untuk menentukan hasil deteksi serangan.

Dan keamanan data pribadi dari si pengguna dan ecommerce bisa diatasi untuk permasalahan menggunakan CNN dari penelitian sebelumnya dikarena menggunakan fitur minimal agar bisa untuk me deteksi suatu serangan seperti brute force lebih cepat dan minimum. Disini peneliti ingin mengetahui kinerja dari DBN

untuk mendeteksi khususnya di Brute force yang saya fokuskan didalam penelitian ini.

5.2. Saran

Peneliti menyarankan menggunakan deep learning yang lebih terbaru lagi untuk mendeteksi suatu jaringan cyber yang menyerang, dan deep learning yang lebih dari CNN dari hasil penelitian ini. Jadi ketika ada deep learning yang lebih bagus dari CNN untuk mendeteksi suatu serangan maka dalam penelitian tersebut dikatakan berhasil lebih baik dari hasil penelitian ini dan penelitian sebelumnya.

Peneliti dapat berharap agar pada penelitian selanjutnya bisa menggunakan deep learning yang terbaru lagi untuk pencarian datanya. Misalkan seperti Generative Adversial network (GAN). Deep learning yang diuji didalam ini antara lain adalah CNN, k-Nearest Neighbour, Logistic Regression dan Support Vector Machine, dan deep belief network.

DAFTAR PUSTAKA

Agghey, Abel Z. et al. 2021. "Detection of Username Enumeration Attack on Ssh Protocol: Machine Learning Approach." *Symmetry* 13(11): 1–13.

Aytaç, Tuğba, Muhammed Ali Aydın, and Abdül Halim Zaim. 2020. "Detection DDOS Attacks Using Machine Learning Methods." *Electrica* 20(2): 159–67.

Ferrag, Mohamed Amine, Lei Shu, Hamouda Djallel, and Kim Kwang Raymond Choo. 2021. "Deep Learning-Based Intrusion Detection for Distributed Denial of Service Attack in Agriculture 4.0." *Electronics (Switzerland)* 10(11): 1–26.

Highnam, Kate, Kai Arulkumaran, Zachary Hanif, and Nicholas R. Jennings. 2021. "BETH Dataset: Real Cybersecurity Data for Unsupervised Anomaly Detection Research." *CEUR Workshop Proceedings* 3095: 1–12.

Ismail et al. 2022. "A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks." *IEEE Access* 10: 21443–54.

Kareem Kamoona, Karrar Raoof, and Cenk Budayan. 2019. "Implementation of Genetic Algorithm Integrated with the Deep Neural Network for Estimating at Completion Simulation." *Advances in Civil Engineering* 2019.

Manimurugan, S. et al. 2020. "Effective Attack Detection in Internet of Medical Things Smart Environment Using a Deep Belief Neural Network." *IEEE Access* 8: 77396–404.

No, Vol, Sari Sandra, and Ahmad Heryanto. 2016. "Visualisasi Serangan Brute Force Menggunakan Metode K-Means Dan Naïve Bayes." 2(1): 315–20.

Priya, S. 2021. "Performance Analysis Comparison on Various Cyber-Attack Dataset By Relating a Deep Belief Network Model on an Intrusion …." *Information Technology in Industry* 9(3): 608–13. http://www.it-in-industry.org/index.php/itii/article/view/600.

Scholarworks, M. 2019. "Utilizing Machine Learning Classifiers to Identify SSH Brute Force Attacks Utilizing Machine Learning Classifiers to Identify SSH

Brute Force Attacks Advisor : James Deverick."

Sohn, Insoo. 2021. "Deep Belief Network Based Intrusion Detection Techniques: A Survey." *Expert Systems with Applications* 167(October 2020).

Wanjau, Stephen Kahara, Geoffrey Mariga Wambugu, and Gabriel Ndung'u Kamau. 2021. "SSH-Brute Force Attack Detection Model Based on Deep Learning." *International Journal of Computer Applications Technology and Research* 10(01): 42–50.

(Agghey et al. 2021; Aytaç, Aydın, and Zaim 2020; Ferrag et al. 2021; Highnam et al. 2021; Ismail et al. 2022; Priya 2021)(Kareem Kamoona and Budayan 2019; No, Sandra, and Heryanto 2016; Scholarworks 2019)