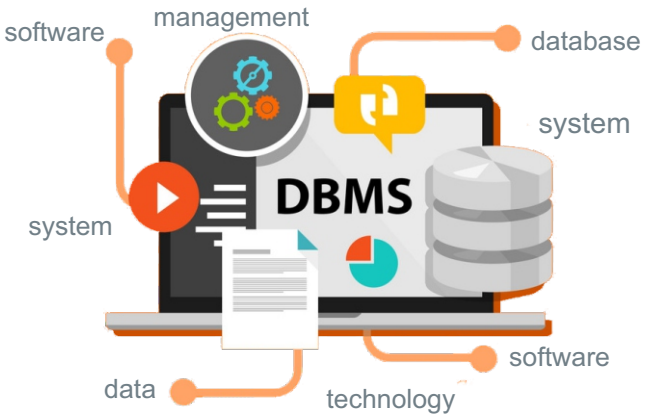
E-ISSN : 2621-0827

ISSN : 2615-7292

JURNAL SIMADA

Sistem Informasi & Manajemen Basis Data



Institut Informatika & Bisnis
DARMAJAYA
Yayasan Alfian Husin

PENGANTAR REDAKSI

Puji Syukur kehadirat allah SWT, atas karunia dan rahmatnya sehingga Jurnal Ilmiah Sistem Informasi dan Manajemen Basis Data (SIMADA) Volume 02, No. 02 bulan Oktober 2019 dapat diterbitkan sesuai dengan periode yang telah ditetapkan.

Jurnal Sistem Informasi dan Manajemen Basis Data (SIMADA) merupakan Jurnal yang diterbitkan oleh Jurusan Sistem Informasi Institut Informatika dan Bisnis (IIB) Darmajaya. Penerbitan jurnal ini sebagai wadah informasi berupa hasil penelitian, studi kepustakaan, gagasan, aplikasi teori dan kajian analisis kritis di bidang keilmuan Sistem Informasi dan Manajemen Basis Data.

Pada edisi ini terdapat 8 artikel dimana versi *online* dari Jurnal tersebut dapat dilihat di jurnal.darmajaya.ac.id. Kami ucapkan terima kasih banyak kepada semua pihak yang telah memberikan kontribusi dalam volume jurnal ini. Pada kesempatan ini kami kembali mengundang dan memberikan kesempatan kepada para peneliti, dibidang Sistem Informasi dan Manajemen Basis Data untuk kembali mempercayai jurnal SIMADA sebagai wadah bagi para peneliti dalam mempublikasikan hasil penelitiannya dalam jurnal ini.

Akhir kata redaksi berharap agar makalah dalam jurnal ini dapat memberikan kontribusi dan sumbangsih pemikiran yang bermanfaat dalam menjawab tantangan yang dihadapi khususnya bagi perkembangan ilmu dan teknologi dalam bidang Sistem Informasi dan Manajemen Basis Data.

Bandar Lampung, Oktober 2019

Redaksi Jurnal Simada

IMPLEMENTASI DATA MINING DENGAN ALGORITMA BERBASIS TREE UNTUK KLASIFIKASI SERANGAN PADA INTRUSION DETECTION SYSTEM (IDS)

Agusa Navirgo[1], Ahmad Habibbullaah[2]

[2]Sekolah Global Surya

[1]*agus.navirgo.1821210003@mail.darmajaya.ac.id*
[2]*ahmad.habibullaah.1821210004@mail.darmajaya.ac.id*

Abstract

Saat ini banyak sekali layanan publik dan komersial yang digunakan melalui Internet, sehingga keamanan sistem menjadi isu terpenting di masyarakat dan ancaman dari peretas juga meningkat. Begitu banyak peneliti merasa sistem deteksi intrusi bisa menjadi garis pertahanan mendasar. Intrusion Detection System (IDS) merupakan sebuah kemampuan yang dimiliki oleh sebuah sistem atau perangkat untuk dapat melakukan deteksi terhadap serangan yang mungkin terjadi dalam jaringan baik lokal maupun yang terhubung dengan internet. Masalah dimulai ketika paket data yang datang sangat banyak dan harus di analisa di kemudian hari. Data mining adalah salah satu solusi mengatasi permasalahan IDS. Makalah ini mengusulkan penggunaan dataset KDDCUP'99 sebagai pengujian awal untuk menganalisis algoritma data mining pada klasifikasi serangan. Algoritma data mining yang diusulkan adalah yang berbasis Tree yaitu Hoeffding Tree, J48, Random Forest, Random Tree dan Rep Tree, kemudian dilakukan pengujian dengan Weka Tools. Hasil yang didapatkan dengan metode 10 fold cross validation pada algoritma Random Forest menghasilkan akurasi tertinggi mencapai 99,9891 %.

Keywords: *Data Mining; Intrusion Detection System; Decision Tree J48; Hoeffding Tree; Rep Tree; Random Forest; Random Tree; KDD dataset*

Abstrak

Saat ini banyak sekali layanan publik dan komersial yang digunakan melalui Internet, sehingga keamanan sistem menjadi isu terpenting di masyarakat dan ancaman dari peretas juga meningkat. Begitu banyak peneliti merasa sistem deteksi intrusi bisa menjadi garis pertahanan mendasar. Intrusion Detection System (IDS) merupakan sebuah kemampuan yang dimiliki oleh sebuah sistem atau perangkat untuk dapat melakukan deteksi terhadap serangan yang mungkin terjadi dalam jaringan baik lokal maupun yang terhubung dengan internet. Masalah dimulai ketika paket data yang datang sangat banyak dan harus di analisa di kemudian hari. Data mining adalah salah satu solusi mengatasi permasalahan IDS. Makalah ini mengusulkan penggunaan dataset KDDCUP'99 sebagai pengujian awal untuk menganalisis algoritma data mining pada klasifikasi serangan. Algoritma data mining yang diusulkan adalah yang berbasis Tree yaitu Hoeffding Tree, J48, Random Forest, Random Tree dan Rep Tree, kemudian dilakukan pengujian dengan Weka Tools. Hasil yang didapatkan dengan metode 10 fold cross validation pada algoritma Random Forest menghasilkan akurasi tertinggi mencapai 99,9891 %.

Kata Kunci: *Data Mining*; *Intrusion Detection System*; *Decision Tree J48*; *Hoeffding Tree; Rep Tree; Random Forest; Random Tree; KDD dataset*

1. PENDAHULUAN

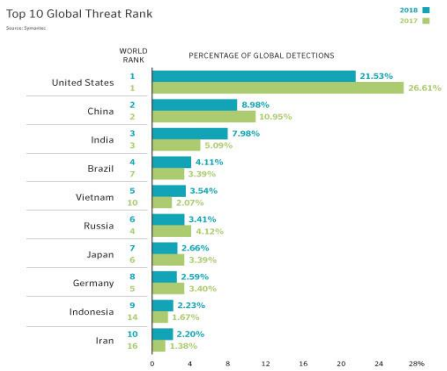
Keamanan sistem atau jaringan menjadi hal yang sangat penting saat ini, khususnya dalam pengamanan informasi yang terdapat dalam sistem atau jaringan tersebut. Informasi memiliki sifat *integrity*, *availability* (ketersediaan), dan *confidentiality* (kerahasiaan). Informasi bagi sebuah perusahaan adalah modal yang sangat penting dan jika salah satu satu dari sifat tersebut terganggu, maka keamanan sistem atau jaringan dari perusahaan tersebut harus segera dilakukan perbaikan.

Menurut G. J. Simons, keamanan sistem informasi adalah bagaimana kita dapat mencegah penipuan (*cheating*) atau, paling tidak, mendeteksi adanya penipuan di sebuah sistem yang berbasis informasi, dimana informasinya sendiri tidak memiliki arti fisik. Selain itu keamanan sistem informasi bisa diartikan sebagai kebijakan, prosedur, dan pengukuran teknis yang digunakan untuk mencegah akses yang tidak sah, perubahan program, pencurian, atau kerusakan fisik terhadap sistem informasi. Hal penting di kehidupan sehari hari yang harus diperhatikan dalam keamanan sistem informasi dan jaringan komputer yaitu kehilangan data/*data loss* dan penyusup/*intruder*.

Kerusakan pada sistem informasi mengakibatkan data tidak dapat diakses atau bahkan hilang dan hal tersebut dapat terjadi setiap saat. Ada banyak hal yang dapat menyebabkan kerusakan tersebut terjadi, diantaranya bencana, *maintenance* (perawatan), kesalahan perangkat lunak, *hardware* (perangkat keras) dan *human error* (kesalahan manusia). Membuat *system backup* dan *recovery data* dapat meminimalisir kehilangan data/*data loss*.

Menurut Bace dan Mell, penyusupan/*intrusion* adalah kegiatan yang merusak atau menyalahgunakan sistem atau setiap usaha yang melakukan *compromise integritas* kepercayaan atau ketersediaan suatu sumber daya komputer dan tidak bertanggung pada berhasil atau tidaknya aksi tersebut sehigga ini berkaitan dengan suatu serangan pada sistem komputer.

Berdasarkan data yang dirilis oleh Symantec pada Internet Security Threat Report tahun 2019 Indonesia masuk peringkat ke-9 dari 157 negara yang terdeteksi mendapat ancaman kejahatan siber terbanyak pada 2018. Ranking Indonesia ini naik dibandingkan tahun sebelumnya, yaitu urutan ke-14 dari 157 negara



Gambar 1. Peringkat Ancaman Kejahatan Siber di Dunia (Top Ten)

Top 10 Threats	Trend since 2016
Infiltration of Malware via Removable Media and External Hardware	↗
Malware Infection via Internet and Intranet	↗
Human Error and Sabotage	↑
Compromising of Extranet and Cloud Components	↑
Social Engineering and Phishing	↘
(D)Dos Attacks	↑
Control Components Connected to the Internet	→
Intrusion via Remote Access	→
Technical Malfunctions and Force Majeure	↘
Compromising of Smartphones in the Production Environment	→

Gambar 2. Ancaman dan serangan pada ICS (publikasi BSI – Federal Office for Information Security)

Penggunaan *Intrusion Detection System* (IDS) yang digunakan bersama dengan *firewall* menjadi standar keamanan sistem dan jaringan. IDS ada bukan untuk menggantikan *firewall*, begitu juga sebaliknya. IDS yang telah ada memiliki

keterbatasan dalam kemampuan beradaptasi dengan sejumlah besar data dan jenis serangan baru. Sehingga muncul dugaan bahwa IDS yang ada sudah hampir tidak dapat mendeteksi semua serangan serangan yang berbahaya yang dilakukan dengan teknik yang baru, tersembunyi atau keduanya dengan jumlah yang besar. Permasalahan ini menyebabkan diperlukan sebuah sistem yang dapat membantu analis dalam proses analisis data dan dapat menemukan serangan yang tidak dapat ditemukan oleh analis atau sensor.

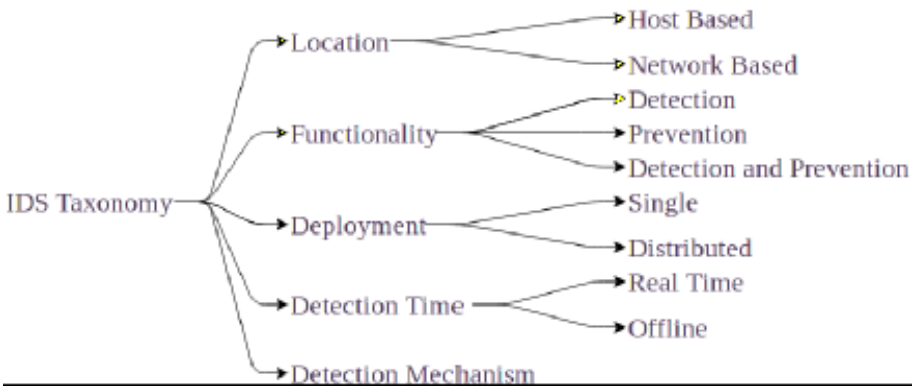
Deteksi intrusi bertujuan untuk menginspeksi dan menemukan gangguan ke sumber daya informasi, dengan melakukan pengamatan, analisis dan mencari bukti berbagai kegiatan percobaan intrusi di sistem dan jaringan. Tiga metoda *Intrusion Detection System* (IDS) berdasarkan bagaimana cara untuk mendeteksi serangan tersebut, yaitu berbasis aturan (*rule based/signature based detection*) atau *misuse detection* , berbasis anomali (*anomaly based detection*) dan *stateful protocol analysis*.

Penelitian mengenai *Intrusion Detection System* (IDS) sudah dimulai sejak tahun 1980 hingga saat ini masih berlanjut untuk mencari metoda *intrusion detection* yang lebih baik dalam kinerja dan performanya. Untuk mengatasi masalah keterbatasan ini, H.Liao et al dalam survei tentang IDS menilai perlu adanya pengetahuan dan terobosan baru untuk memperbaiki keterbatasan teknik, performa dan kinerja IDS yang ada sekarang. Sugiantoro dalam penelitian nya untuk sistem deteksi penyusupan merekomendasikan teknik *mobile agent* sebagai teknologi baru dalam *Intrusion Detection System* secara NIDS, tapi masih perlu penelitian lebih lanjut untuk menerapkan secara global model *mobile agent* ini.

Data mining adalah salah satu solusi mengatasi permasalahan IDS. Banyak masalah baru yang muncul telah dipecahkan dengan metoda *data mining* seperti masalah statistik, algoritme komputasi, teknologi *database*, komputasi tingkat tinggi, *machine learning*, pengenalan pola, dan sebagainya. Masalah pada *network setting* menjadi salah satu tantangan dalam *data mining*.

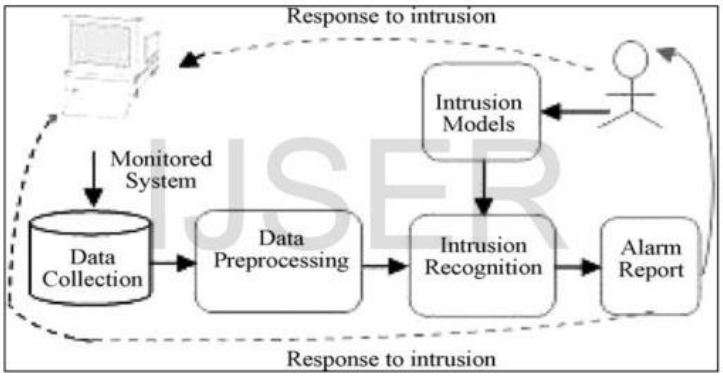
2. KERANGKA TEORI

IDS *(Intrusion Detection System)* merupakan sistem untuk mendeteksi adanya "intrusion" yang dilakukan oleh "intruder" atau "pengganggu atau penyusup" di jaringan. IDS *(Intrusion Detection System)* sangat mirip seperti alarm, yaitu IDS (Instrusion Detection System) akan memperingati bila terjadinya atau adanya penyusupan pada jaringan. IDS *(Intrusion Detection System)* dapat didefinisikan sebagai kegiatan yang bersifat anomaly, incorrect, inappropriate yang terjadi di jaringan atau host. IDS *(Intrusion Detection System)* adalah sistem keamanan yang bekerja bersama Firewall untuk mengatasi Intrusion.



Gambar 3. Taksonomi IDS

Taksonomi pada IDS dibagi menjadi lima bagian. Setiap bagian tersebut haru diperhitungkan berdasarkan tujua penggunaannya dan keuntungan kerugianny Kelima hal tersebut adalah lokasi, fungs penyebarannya, waktu pendeteksiannya da mekanisme pendeteksiannya (Pharate, dkk., 2015). Gambar 2 menjelaskan tentang taksonomi dari IDS.



Gambar 4. Tahapan Membangun IDS

Untuk membangun IDS setidaknya diperlukan 5 (lima) tahapan yaitu Data Collection, Data Preprocessing, Intrusion Recognition, Reporting dan Respone. IDS () juga memiliki cara kerja dalam menganalisa apakah paket data yang dianggap sebagai intrusion oleh intruser. Cara kerja IDS () dibagi menjadi dua, yaitu :

a. Knowledge Based (Misuse Detection)

Knowledge Based pada IDS *(Intrusion Detection System)* adalah cara kerja IDS *(Intrusion Detection System)* dengan mengenali adanya penyusupan dengan cara menyadap paket data kemudian membandingkannya dengan database rule pada IDS *(Intrusion Detection System)* tersebut. Database rule tersebut dapat berisi signature – signature paket serangan. Jika pattern atau pola paket data tersebut terdapat kesamaan dengan rule pada database rule pada IDS *(Intrusion Detection System)*, maka paket data tersebut dianggap sebagai seranganm dan demikian juga sebaliknya, jika paket data tersebut tidak memiliki kesamaan dengan rule pada database rule pada IDS*(Intrusion Detection System)*, maka paket data tersebut tidak akan dianggap serangan.

b. Behavior Based (Anomaly Based)

Behavior Base adalah cara kerja IDS *(Intrusion Detection System)* dengan mendeteksi adanya penyusupan dengan mengamati adanya kejanggalan – kejanggalan pada sistem, aatu adanya keanehan dan kejanggalan dari kondiri pada saat sistem normal, sebagai: adanya penggunaan memory yang melonjak secara terus menerus atau terdapatnya koneksi secara paralel dari satu IP dalam jumlah banyak dan dalam waktu yang bersamaan. Kondisi tersebut dianggap kejanggalan yang selanjutnya oleh IDS *(Intrusion Detection System) Anomaly Based* ini dianggap sebagai serangan. Intrusion itu sendiri didefinisikan sebagai kegiatan yang bersifat anomaly, incorrect, inappropite yang terjadi di jaringan atau di host tersebut. Intrusion tersebut kemudian akan diubah menjadi "*rules*" ke dalam IDS *(Intrusion Detection System)*. Sebagai contoh, *intrusion* atau gangguan seperti *port scanning* yang dilakukan oleh intruder. Oleh karena itu IDS *(Intrusion Detection System)* ditujukan untuk meminimalkan kerugian yang dapat ditimbulkan dari intrusion. Kelebihan yang akan di dapatkan dengan menggunakan IDS *(Intrusion Detection System)* sebagai metode Keamanan:

1. Memiliki Akurasi keamanan yang baik

   IDS *(Intrusion Detection System)* haruslah memiliki akurasi atau ketelitian, jadi IDS *(Intrusion Detection System)* yang baik adalah IDS *(Intrusion Detection System)* yang memiliki ketelitian yang baik untuk mengenal intrusion

atau gangguan. Pada saat sekrarang ini IDS *(Intrusion Detection System)* telah memiliki ketelitian tinggi, yaitu mampu secara realtime mendeteksi dan melakukan blocking terhadap tindakan yang mencurigakan. Selain itu IDS () juga harus mampu memeriksa dan menganalisa pattern objek secara menyeluruh seperti paket – paket data baik Header Paket maupun Payload yang dipergunakan serta membedakan paket data yang keluar masuk dalam lalu lintas jaringan sehingga dapat mengenal benar karateristik trafic.

2. Mampu Mendeteksi dan Mencegah Serangan.

   IDS *(Intrusion Detection System)* haruslah dapat mendeteksi serangan dn juga mampu untuk melakukan pencegahan terhadap serangan tersebut, IDS *(Intrusion Detection System)* yang baik dalam mengatasi serangan adalah IDS *(Intrusion Detection System)* yang memiliki karateristik:

   - Dapat beroperasi secara *in-line*.
   - Memiliki kehandalan dan ketersediaan.
   - Deliver high performance.
   - Kebijakan policy pada IDS *(Intrusion Detection System)* yang dapat diatus sesuai dengan yang dibutuhkan.

3. Memiliki cakupan yang Luas dalam Mengenal Proses *Attacking*

   IDS *(Intrusion Detection System)* haruslah memiliki pengetahuan yang luas, dapat mengenal serangan apa yang belum dikenalnya, seperti contoh IDS*(Intrusion Detection System)* harus mampu mendeteksi serangan DOS mempergunakan analisis signature dan mampu mendeteksi segala sesuatu yang mencurigakan. IDS *(Intrusion Detection System)* yang baik dalam pengenalan attacking adalah IDS *(Intrusion Detection System)* yang memiliki karateristik:

   - Memiliki AI () sehingga IDS *(Intrusion Detection System)* tersebut dapat mempelajari sendiri serangan – serangan yang datang.
   - Mampu melakukan proses deteksi trafic dan pembersihan terhadap host (Layer 3 – Layer 7).
   - Mampu melakukan scanning TCP dan UDP.
   - Mampu memeriksa keberadaan backdoor.

4. Dapat memeberikan Informasi tentang ancaman – ancaman yang terjadi.
5. Memiliki tingkat Forensik yang canggih dan mampu menghasilkan reporing yang baik. Memiliki sensor yang dapat dipercaya untuk memastikan pendeteksian dan pencegahan.

2.1 Data Mining

Data mining adalah teknologi yang mengombinasikan metode analisis tradisional dan algoritma yang canggih agar proses data besar lebih cepat diproses. *Data mining* biasa disebut dengan sebutan yang sering digunakan untuk mencari pengetahuan yang tersembunyi didalam *database*. *Data mining* menggunakan teknik statistika, matematika, kecerdasan buatan, dan *machine learining* untuk mengekstraksi dan menganalisa informasi yang terdapat dalam *database* besar. (Turban et al, 2005).

Analisis yang dilakukan *data mining* lebih maksimum dibandingkan sistem pendukung keputusan tradisional yang banyak digunakan. *Data mining* mengatasi masalah-masalah bisnis dengan cara tradisional yang menggunakan banyak waktu da biaya yang tinggi. *Data mining* menjelajahi basis data untuk mengetahui pola yang tersembunyi, serta mencari informasi agar dapat memprediksi yang bisa saj dilupakan oleh pembisnis karena kemungkinan besar mereka tidak menduganya.

Pada perkembangan teknologi saat ini, proses pengumpulan data serta penyimpanannya telah mudah dijalankan walaupun data tersebut berukuran besa sehingga *data mining* melakukan proses pencarian secara mudah dan otomati

mencari informasi yang berguna dalam penyimpanan data yang mempunyai ukura yang besar. Istilah ini biasa disebut dengan *Knowledge Discovery in Database* (KDD) yang digunakan secara bergantian untuk memberikan penjelasan tentang prose pencarian informasi yang tersembunyi dalam suatu basis data yang besar. Sebenarnya konsep ini berkaitan satu sama lain walaupun konsepnya berbeda.

2.2 Teknik Data Mining

Data Mining merupakan proses untuk mencari nilai tambah dari beberpa data yang tidak bisa dilakukan secara manual, *Data Mining* menganalisa secara otomatis dari data yang berukuran sangat besar dengan fungsi untuk menemukan pola yang sangat penting terkadang tidak diketahui keberadaannya, dan *Data Mining* dapat memilih informasi yang bermanfaat dari database yang tersembunyi yang sebelumnya tidak dikenali. Proses ini mendekati teknis yang berbeda seperti *Clustering, Data Summarization, Learning Classification, Rules.* Tidak semua proses disebut *Data Mining* dalam mencari informasi, misalnya pencarian rekaman individu menggunakan *database management system* atau pencarian *web* yang menggunakan *query* kesemua *search engine* yang berkaitan dengan information *retrieval* dan *data mining* digunakan untuk meningkatkan kemampuannya.

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Perlu diingat bahwa kata *mining* sendiri berarti usaha untuk mendapatkan sedikit data berharga dari sejumlah besar data dasar. Karena itu *data mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan *(artificial intelligent), machine learning,* statistik dan basisdata. Beberapa teknik yang sering disebut-sebut dalam literatur *data mining* antara lain yaitu *association rule mining, clustering, klasifikasi, neural network, genetic algorithm* dan lain-lain.

Klasifikasi biasanya berhubungan dengan peramalan kategori kelas dan menggolongkan data atau membangun sebuah model yang berdasarkan dengan pelatihan data untuk menetapkan dan nilai-nilai kelas dalam sebuah golongan atribut dan menggunakan golongan data baru. Klasifikasi sering digunakan dalam bidang persetujuan kredit, target marketing, diagnose medis, dan analisa keefektifan sebuah keputusan. Langkah klasifikasi dengan menguraikan sebuah himpunan kelas yang telah ditentukan dan menggunakan model yang berfungsi untuk mengklasifikasi tuple data yang label kelasnya belum diketahui. Model-model tersebut disajikan sebagai kaidah klasifikasi, pohon keputusan, atau rumus matematis. Macam-macam klasifikasi yang sering digunakan adalah *Decision Tree, Bayesian Network, Adaptive Bayesian Network, Naïve Bayes, Random Forest, Random Tree* dan lain sebagainya.

2.3 Tahapan Data Mining

Salah satu tuntutan dari *data mining* ketika diterapkan pada data berskala besar adalah diperlukan metodologi sistematis tidak hanya ketika melakukan analisa saja tetap juga ketika mempersiapkan data dan juga melakukan interpretasi dari hasilnya sehingga dapat menjadi aksi ataupun keputusan yang bermanfaat. *Data mining* seharusnya dipahami sebagai suatu proses, yang memiliki tahapan-tahapan tertentu dan juga ada umpan balik dari setiap tahapan ke tahapan sebelumnya. Pada umumnya proses *data mining* berjalan interaktif karena tidak jarang hasil data mining pad awalnya tidak sesuai dengan harapan analisnya sehingga perlu dilakukan desain ulan prosesnya.

Sesuai yang tercantum dalam buku *"Advances in Knowledge Discovery* dan *Data mining"* terdapat definis i sebagai berikut:*Knowledge discovery (data mining) in databases* (KDD) adalah keseluruhan proses *non-trivial* untuk mencari dan mengidentifikasi pola *(pattern)* dalam data, dimana pola yang ditemukan bersifat sah *(valid),* baru *(novel),* dapat bermanfaat *(potentially usefull),* dapat dimengerti *(ultimately understandable).* (Vapnik, 1998). Istilah *data mining*

dan *Knowledge Discovery In Databases* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbedaakan tetap berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining.* Proses KDD secara garis besar dapat dijelaskan sebagai berikut:

a. Data Selection

Pemilihan (seleksi) data dan sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining,* disimpan dalam suatu berkas, terpisah dari basis data operasional.

b. *Pre-processing / Cleaning*

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment,* yaitu pros es "memperkaya" data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

c. Transformation Coding

Proses transformasipada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining.* Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

d. Data mining

Proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi.

2.4 *Machine Learning*

a. *Machine Learning* dapat digunakan untuk melakukan penggalian informasi pada data set yang tersedia. Dengan menggunakan perhitungan statistika dan algoritma yang matematis, machine learning dapat mengetahui informasi yang tersembunyi, pola dan hubungan antar atribut dalam sebuah data set. Fungsi ini menjadi sangat berguna untuk mengetahui data yang mencurigakan.

b. *Machine Learning* juga dapat digunakan untuk mendeteksi serangan pada jaringan (J. dan Muthukumar, 2015). Pengembangan terhadap penggunaan machine learning telah dikembangkan untuk mengetahui algoritma yang terbaik untuk *detect ion engine* pada IDS Tabel 1 menunjukkan perbandingan performa antar algoritma yang diimplementasikan pada IDS .

Tabel 1. Akurasi Algoritma *Machine Learning*

Algoritma	**Akurasi**
Adaboost	92.2073
Hyperpipes	92.2363
J48	***96.25 74***
Naïve Bayes	90.5504
OneR	94.5741
Random Forest	35.8247
Random Tree	96.225 8
ZeroR	92.2073

2.5 Weka

Weka merupakan suatu perangkat lunak yang berisikan koleksi dari perangkat visualisasi dan algoritma untuk analisis data dan *predictive modelling,* termasuk dengan tampilan antar muka yang mudah diakses oleh pengguna. Produk asli Weka ini sebenarnya adalah TCL/TK yakni sebuah pemodelan algoritma yang diimplementasikan dalam bahasa pemrograman lain, termasuk utiliti pemprosesan data dalam bahasa C, dan *Makefile* sistem untuk dijalankan sebagai eksperimen *machine learning.* Versi perangkat lunak ini pada awalnya dibuat sebagai alat untuk menganalisis data agrikultura, tetapi setelah muncul versi Weka yang dikembangkan sejak dari tahun 1997, maka kini Weka banyak digunakan dalam berbagai area aplikasi yang lainnya, terutama dibidang edukasi dan penelitian.

Beberapa kelebihan utama Weka antara lain:

a. Merupakan perangkat lunak gratis yang dapat disebarluaskan dan digunakan yang memiliki naungan lisensi dibawah GNU General Public License.
b. Bersifat sangat *portable* karena dapat diimplementasikan dalam pemrograman Java dan dapat berjalan diberbagai platform sistem komputer saat ini.
c. Berisikan koleksi yang meliputi berbagai teknik *pre-processing* dan teknik permodelan data.
d. Mudah digunakan oleh pemula karena terdapat antar muka grafis yang mudah dipahami bagi orang awam sekalipun.

Weka juga mendukung berbagai tugas standar untuk *data mining,* lebih spesifik dibidang seperti data *pre-processing, clustering, classification, regression,* visualisasi dan seleksi fitur. Algoritma yang didukung Semua teknik dari Weka diprediksikan berdasarkan asumsi bahwa data adalah sebuah data tunggal yang datar atau relasi, dimana setiap point data dideskripsikan dengan nomor dari atribut. Weka juga mendukung akses ke database SQL menggunakan *Java Database Connectivity* dan dapat memproses hasilnya dikembalikan kebentuk queri *databas*e. Meskipun hal ini bukanlah suatu multi-relational *data mining* tapi ada perangkat lunak terpisah yang mampu mengkonversi koleksi dari tabel *database* yang terhubung kepada sebuah tabel yang cocok untuk diproses melalui Weka. Classifier yang ada di Weka antara lain *Decision Stump, j48, NB Tree, Rep Tree, Random Forest, Random Tre, Hoefding Tree,* dan lain sebagainya.

3. METODOLOGI

Dalam pelaksanaan dan pengimplementasian sistem deteksi serangan menggunakan algoritma berbasis tree penulisa melakukan langkah-langkah sebagai berikut:

a. Survei tentang berbagai metode untuk menangani masalah deteksi intrusi.
b. *Pre-processing*

   Data intr usi yang digunakan untuk percobaan diambil dari dataset KDD CUP'99, yang mana dataset ini sudah menjadi patokan oleh banyak peneliti. "10% dari KDD CUP"dipilih dari KDD CUP'99 dataset untuk mengevaluasi rules dan pengujian data guna mendeteksi intrusi, koneksi diberi label normal atau attack, dikategorikan dalam 4 kelas kategori utama yaitu:

   a. DOS (Denial -of-Service - serangan yang berusaha menggagalkan layanan server), termasuk di dalamnya : Apache2, arppoison back, Crashiis, dosnuke, Land, Mailbomb, SYN Flood, (Neptune), Ping of Death (POD), Process Table, selfping, Smuff.
   b. PROBE (seperti Port Scanning yang berusaha mencari kelemahan sistem yang ada), misal : insidesniffer, Ipsweep, ls_domain, Mscan, NTinfoscan, Nmap, queso, resetscan, Saint, Satan.

c. U2R (*unauthorized access to root privileges*) yang melakukan akses yang bukan haknya ke superuser dari jaringan dalam), termasuk dalam kategori ini : anypw, casesen, Eject, Ffbconfig, Fdformat, Loadmodule, ntfsdos, Perl, Ps, sechole, Xterm, yaga.

d. R2L (*unauthorized remote login to machine*) yang melakukan akses yang tidak bukan haknya dari jarak jauh), termasuk dalam kategori ini: Dictionary, Ftpwrite, Guest, Httptunnel, Imap, Named, ncftp, netbus, netcat, Phf, ppmacro, Sendmail, sshtrojan, Xlock, Xsnoop.

Tabel 2. Kelas dan Serangan

Kelas	Serangan
DOS	apache, back, land, mailbomb, neptune, pod, processtable, smurf, teardrop, udpstorm
PROBE	ipsweep, mscan, nmap, portsweep, saint, satan
U2R	buffer_overflow, loadmodule, perl, rootkit, ps, sqlattack, xterm
R2L	ftp_write, guess_password, imap, multi hop

Dataset KDD CUP'99 tersedia pada dengan total data 494.021 record secara detail ditunjukkan pada Tabel 3 berikut ini.

Tabel 3. Rincian Distribusi Kelas dan Serangan

Serangan	Jumlah record	Kelas	Jumlah record tiap kelas
Back	2203	DOS	
Land	21	DOS	
Neptune	107201	DOS	391458
Pod	264	DOS	
Smurf	280790	DOS	
Teardrop	979	DOS	
Satan	1589	PROBE	
Ipsweep	1247	PROBE	
Nmap	231	PROBE	4107
Portsweep	1040	PROBE	
Normal	97278	NORMAL	97278
Guess_passwd	53	R2L	
ftp_write	8	R2L	
Imap	12	R2L	
Phf	4	R2L	
Multihop	7	R2L	1126
Warezmaster	20	R2L	
Warezclient	1020	R2L	
Spy	2	R2L	
Buffer_overflow	30	U2R	
Loadmodule	9	U2R	
Perl	3	U2R	52
Rootkit	10	U2R	

Pada dataset KDD CUP terdapat 1 data normal data dan 22 jenis serangan yang dikelompokkan kedalam 4 kategori serangan yaitu DOS, Probe, R2L, dan U2R. Dalam penelitian ini penulis hanya menggunakan 247.010 record dari 494.021. Pemilihan record dilakukan dengan cara menerapkan *remove percentage* 50 proses pada weka. Distribusi data yang terpilih ditunjukkan pada tabel 4 berikut ini.

Tabel 4. Rincian Distribusi Kelas dan Serangan

Label Kelas	Jumlah record
Dos	220145
Probe	788
Normal	26053
R2L	1
U2R	23

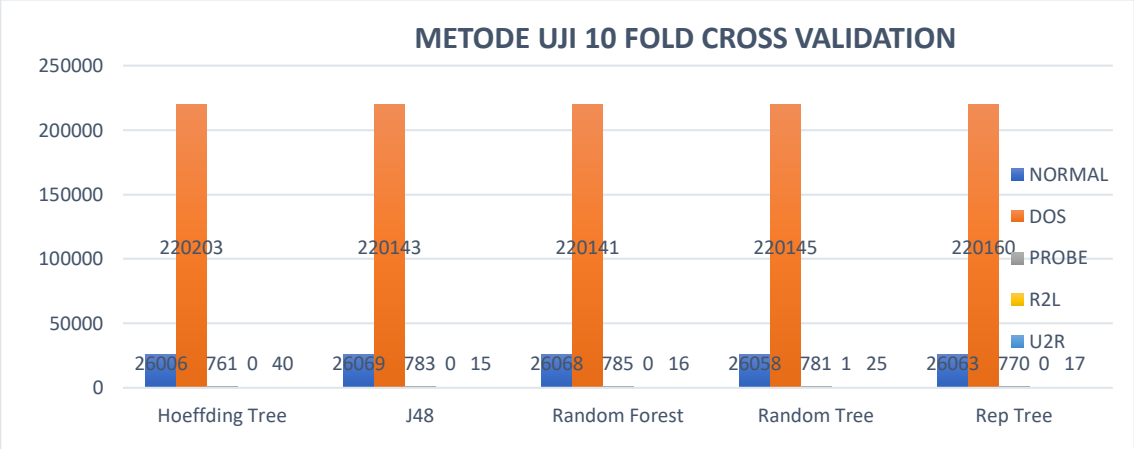
Eksperimen dilakukan pada Hardware HP EliteOne 800 G2 23-in Touch AiO Intel Core ™ I7 677 CPU @ 3,41 GHz 3.41 GHz RAM 4 GB dan Software Windows 10 serta menggunakan WEKA 3.9 unntuk pengolahan dataset. Pengujian dilakukan dengan menggunakan Algoritma berbasis Tree yaitu Hoeffding Tree, J48, Random Forest, Random Tree dan Rep Tree yang diuji pada metode 10 Fold Cross Validation dan Split 66 %.

4. HASIL DAN PEMBAHASAN

Pada bagian ini akan diuraikan hasil eksperimen yang telah dilakukan. Hasil uji pada metode 10 Fold Cross Validation terhadap kelas label ditunjukkan pada tabel 6 dan grafik 1 berikut ini.

Tabel 5. Hasil Uji Dengan Metode 10 Fold Cross Validation

Algoritma Klasifikasi	Jenis Serangan				
	NORMAL	DOS	PROBE	R2L	U2R
Hoeffding Tree	26006	220203	761	0	40
J48	26069	220143	783	0	15
Random Forest	26068	220141	785	0	16
Random Tree	26058	220145	781	1	25
Rep Tree	26063	220160	770	0	17

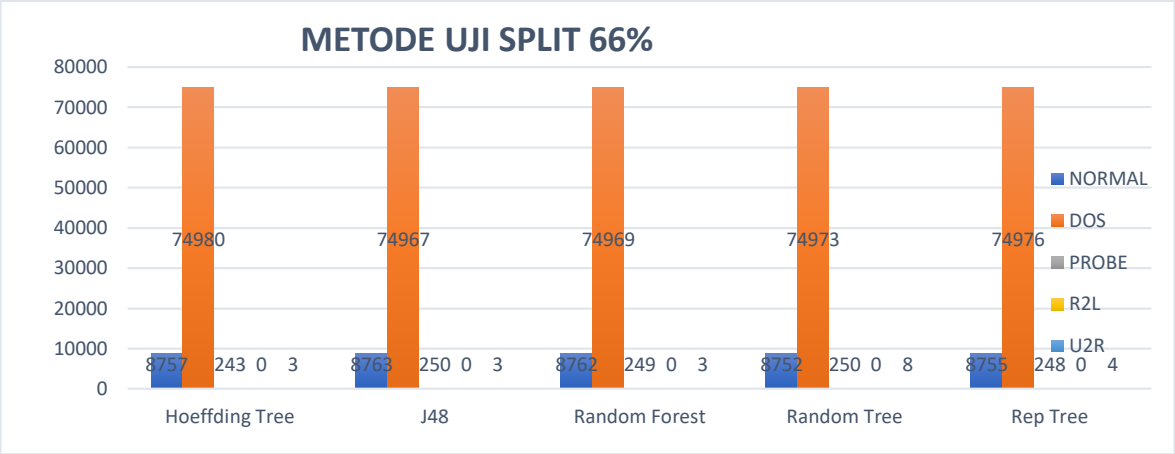


Grafik 1. Hasil Uji Dengan Metode 10 Fold Cross Validation

Kemudian dilanjutkan dengan metode split 66% terhadap kelas label dengan hasil pada tabel 6 dan grafik 2 berikut ini:

Tabel 6. Hasil Uji Dengan Metode Split 66%

Algoritma Klasifikasi	Jenis Serangan				
	NORMAL	DOS	PROBE	R2L	U2R
Hoeffding Tree	8757	74980	243	0	3
J48	8763	74967	250	0	3
Random Forest	8762	74969	249	0	3
Random Tree	8752	74973	250	0	8
Rep Tree	8755	74976	248	0	4



Grafik 2. Hasil Uji Dengan Metode Split 66%

Selanjutnya Hasil perbandingan kedua metode uji yang diperoleh berdasarkan output pada tool WEKA ditunjukkan pada Tabel 7 berikut ini:

Tabel 8. Evaluasi Performansi

Algoritma Klasifikasi	CCI		ICI		MAE (%)	RMSE (%)	RAE (%)	Time Build Model (s)	Metode Uji
	Jumlah	%	Jumlah	%					
Hoeffding Tree	83.947	99,9571	36,00	0,0429	0,0005	0,0104	0,7559	61,84	Split 66 %
Hoeffding Tree	246.874	99,9449	136,00	0,0551	0,0005	0,0103	0,6628	62,59	10-fold cross-validation
Random Forest	83.971	99,9857	12	0,0143	-	0,0038	0,0639	100,16	split 66 %
Random Forest	246.983	99,9891	27	0,0109	-	0,0036	0,0545	91,64	10-fold cross-validation
J48	83.963	99,9762	20,00	0,0238	-	0,0057	0,0634	12,44	split 66 %
J48	246.959	99,9794	51,00	0,0206	-	0,0051	0,0528	11,74	10-fold cross-validation
Random Tree	83.966	99,9798	17	0,0202	-	0,0052	0,0381	1,03	split 66 %
Random Tree	246.958	99,9789	52	0,0211	-	0,0052	0,0395	1,08	10-fold cross-validation

Rep Tree	83.962	99,9750	21	0,0250	0,0001	0,0057	0,0786	8,77	split 66 %
Rep Tree	246.950	99,9757	60	0,0243	0,0001	0,0056	0,0770	8,83	10-fold cross-validation

Beberapa metrik evaluasi kinerja yang bisa digunakan untuk analisis kemampuan model deteksi intrusi, namun untuk penelitian ini ditetapkan fungsi evaluasi kinerja digunakan seperti: *Correctly Classified Instances (CCI), Incorrectly Classified Instances (ICI), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan Relative Absolute Error (RAE). Random Forest* mencapai akurasi tertinggi yaitu 99.9891 % yang di uji dengan metode 10 *Fold Cross Validation* dan waktu *build model* 91,64 detik. Keakuratannya sedikit menurun jika di uji pada metode *Split* 66% menjadi 99.9857 % namun waktu komputasi model menjadi naik hingga 100,16 detik yang merupakan waktu build model tertinggi. Kemudian Random Tree dengan tingkat akurasi 99,9798 % yang diuji dengan metode split 66 %. Keakuratannya sedikit menurun namun waktu komputasi model turun hingga 1,03 detik yang merupakan tingkat efisiensi waktu yang paling optimal.

Hoeffding Tree dengan tingkat akurasi 99, 9571 % dan 99, 9449 % menjadi yang terendah, masing-masing diuji dengan metode *Split* 66% dan 10 Fold Cross Validation, menghasilkan waktu *build model* 61, 84 detik dan 62,59 detik. Tingkat false positif terendah dicapai oleh Hoeffding Tree yang diuji pada 2 (dua) metode *Split* 66% dan 10 *Fold Cross Validation*, ini karena *Hoeffding Tree* berupaya mengoptimalkan margin antara kelas negatif dan inti kelas positif. Tingkat kesalahan pada tiap-tiap eksperimen adalah sangat rendah, hal ini dapat dilihat dari MAE sudah pada angka 0 untuk Random Forest, J48 dan *Random Tree*. Waktu build model paling optimal pada *Random Tree dan Rep Tree* berada pada kisaran kurang dari 10 detik yakni berada di kisaran angka 1 detik dan 8-9 detik.

5. KESIMPULAN

5.1 Kesimpulan

Dalam penelitian ini, dikembangkan lima model untuk memecahkan masalah deteksi intrusi Menggunakan algoritma Hoeffding Tree, Random Tree, J48, Random Forest dan Rep Tree untuk klasifikasi serangan. Penerapan metode uji pada Split 66 % dan 10 Fold Cross Validation dapat meningkatkan tingkat akurasi meskipun sangat kecil yakni pada kisaran 0,0007 s/d 0,0122% untuk klasifikasi serangan dengan menggunakan dataset intrusi KDD CUP'99. Algoritma Random Forest rata-rata memiliki tingkat akurasi tertinggi baik eksperimen dengan metode Split 66 % dan 10 *Fold Cross Validation* yaitu 99, 9857% dan 99,9891%. Sedangkan waktu build model tersingkat pada Random Tree baik eksperimen dengan metode Split 66 % dan 10 Fold Cross Validation dengan waktu 1,03 detik dan 1,08 detik.

5.2 Saran

Agar kelima algoritma yakni *Hoeffding Tree, J48, Random Forest, Random Tree dan Rep Tree* dapat dieksperimen dengan *software* lain seperti Rapidminer yang dikombinasikan dengan pemilihan fitur (fitur *selection*) untuk memperoleh tingkat akurasi yang sangat tinggi dan waktu build model tersingkat.

DAFTAR PUSTAKA

Aggarwala, Preeti., Sharmab, Sudhir Kumar., 2015. Analysis of KDD Dataset Attributes - Class wise For Intrusion Detection, 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015).

Bilal, Ahmad., 2017. Intrusion Detection With Tree-Based Data Mining Classification Techniques By Using Kdd Dataset, European Jurnal Of Computer Science and Information Technology. Vol. 5. No. 6, pp. 11-18, December 2017.

Chandolikar, P. N. S.., Nandavadekar, P. Dr. V.D., 2012. Selection of Relevant Feature for Intrusion Attack Classification by Analyzing KDD Cup 99, MIT Int. J. Comput. Sci. Inf. Technol., vol. 2, no. 2, pp. 85–90, 2012.

Federal Office for Information Security (BSI), 2019. Industrial Control System Security Top 10 Threats and Countermeasures 2019.

J., J. & Muthukumar, D. B., 2015. Intrusion Detection System (ID S): Anomaly Detection Using Outlier Detection Approach. Procedia Computer Science, Volume 48, pp. 338-346.

KDD Cup 1999 Data." [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. [Accessed: 02-Agustus-2019].

Lakshmi Devasena C., 2014. Comparative Analysis of Random Forest, REP Tree and J48 Classifiers for Credit Risk Prediction, International Journal of Computer Applications (0975 – 8887), (ICCCMIT-2014), 30-36.

Lee, W., S, Stolfo., 1998. Data Mining Approaches for Intrusion Detection. 7th USENIX Security Symposium, San Antonio, TX.

Modi, Urvashi., Jain, Prof. Anurag., 2015. A survey of IDS classification using KDD CUP 99 dataset in WEKA, International Journal of Scientific & Engineering Research, Volume 6, Issue 11, November-2015.

R. Bace., P. Mell., 2001. Intrusion Detection System, NIST Spec. Publ. Intrusion Detect. Syst., pp. 1–51, 2001.

Symantec Corporation., 2019. Internet Security Threat Report 2019 Appendices, Internet Secur. Threat Rep., vol. 24, February, 2019.

Turban, Efraim., E Jay., Aronson., Liang Ting-Peng., 2005. *Decision Support System and Intelligent System.* Andi Offset.

Weka machine learning tool available on http://www.cs.waikato.ac.nz/ml/weka/downloading.html.

Witten, I. et al., 1999. Weka: practical machine learning tools and techniques with java implementations.

Diterbitkan :
LEMBAGA PENGEMBANGAN PEMBELAJARAN, PENELITIAN, DAN PENGABDIAN MASYARAKAT (LP4M)
INSTITUT INFORMATIKA & BISNIS DARMAJAYA

Alamat : Jalan Zainal Abidin Pagar Alam No.93 Gedong Meneng, Bandar Lampung 35142
Telp. **0721-787214** Fax. **0721- 700261**
email : simada@darmajaya.ac.id
Website : jurnal.darmajaya.ac.id