Implementasi Data Mining Dengan Algoritma Berbasis Tree Untuk Klasifikasi Serangan Pada Intrusion Detection System (Ids)

Agusa Navirgo [1)], Ahmad Habibbullaah[2)]
[12)]Magister Teknik Informatika - Fakultas Ilmu Komputer
Institut Informatika & Bisnis Darmajaya
Jl. Z.A Pagar Alam No.93 Bandar Lampung Indonesia 35142
Telp: (0721)-787214 Fax (0721)-700261 ext 112
Email: agus.navirgo.1821210003@mail.darmajaya.ac.id [1],
ahmad.habibullaah.1821210004@mail.darmajaya.ac.id[2]

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*

1. Pendahuluan

Keamanan sistem atau jaringan menjadi hal yang sangat penting saat ini, khususnya dalam pengamanan informasi yang terdapat dalam sistem atau jaringan tersebut [3]. Informasi memiliki sifat *integrity*, *availability* (ketersediaan), dan *confidentiality* (kerahasiaan). Informasi bagi sebuah perusahaan adalah modal yang sangat penting dan jika salah satu dari sifat tersebut terganggu, maka keamanan sistem atau jaringan dari perusahaan tersebut harus segera dilakukan perbaikan.

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*. 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 sehingga ini berkaitan dengan suatu serangan pada sistem computer.[1]

Penggunaan *Intrusion Detection System* (IDS) yang digunakan bersama dengan *firewall* menjadi standar keamanan sistem dan 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 ciber terbanyak pada 2018 [4]. Ranking Indonesia ini naik dibandingkan tahun sebelumnya, yaitu urutan ke-14 dari 157 negara. [2]

Deteksi intrusi bertujuan untuk menginspeksi dan menemukan gangguan ke sumber daya informasi, dengan melakukan pengamatan, analisis dan mencari bukti berbagai kegiatan percobaan intruksi 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.*

2. Metode Penelitian

2.1 Metode Pengumpulan Data

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

a. Survei tentang berbagai metode untuk menangani masalah deteksi intrusi.

b. Pre-proses

Data intruksi 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.

Pada tabel 1 berikut dijelaskan beberapa kelas dan serangan.

Tabel 1. Kelas dan Serangan

Kelas	**Serangan**
DOS	apache, back, land, mailbomb, neptune, pod,
PROBE	ipsweep, mscan, nmap,
U2R	buffer_overflow, loadmodule, perl, rootkit, ps,
R2L	ftp_write, guess_password,

Dataset KDD CUP'99 tersedia pada dengan total data 494.021 record secara detail ditunjukkan pada Tabel 2 berikut ini.

Serangan	**Jumlah record**	**Kelas**	**Jumlah record tiap kelas**
Back	2203	DOS	391458
Land	21	DOS	
Neptune	10720	DOS	
Pod	264	DOS	
Smurf	28079	DOS	
Teardrop	979	DOS	
Satan	1589	PROBE	
Ipsweep	1247	PROBE	
Nmap	231	PROBE	
Portsweep	1040	PROBE	
Normal	97278	NORMAL	97278
Guess_passwd	53	R2L	
ftp_write	8	R2L	
Imap	12	R2L	
Phf	4	R2L	
Multihop	7	R2L	
Warezmaster	20	R2L	
Warezclient	1020	R2L	
Spy	2	R2L	
Buffer_overflow	30	U2R	

Loadmodule	9	U2R
Perl	3	U2R
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 3 berikut ini.

Tabel 3. 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 ™ I7677 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 %.

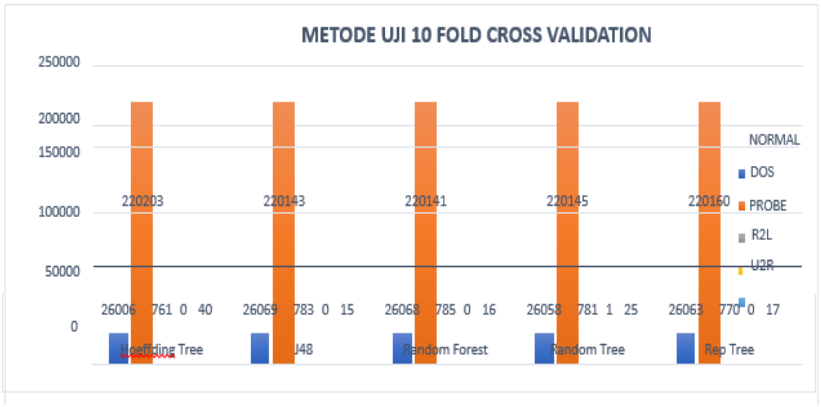
3. HASIL DAN PEMBAHASAN

Pada bagian ini akan diuraikan hasil eksperimen yang telah dilakukan. Hasil uji pada metode 10 Fold Cross Validation terhadap kelas label ditunjukkan pada tabel 4 berikut ini.

Tabel 4. 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

Hasil Uji Dengan Metode 10 Fold Cross Validation terlihat pada grafik 1 berikut ini.

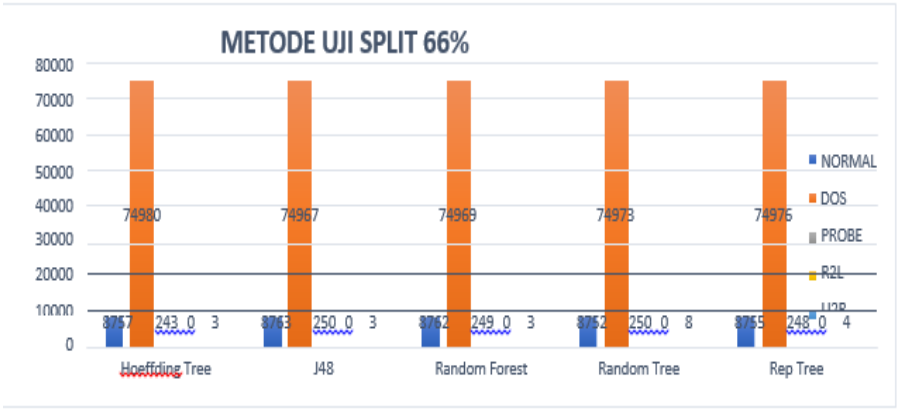


Gambar 1. Grafik Hasil Uji Dengan Metode 10 Fold Cross Validation

Kemudian dilanjutkan dengan metode split 66% terhadap kelas label dengan hasil pada tabel 5 dan grafik 2 berikut ini

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

Tabel 5. Hasil Uji Dengan Metode Split 66%



Gambar 2. Grafik Hasil Uji Dengan Metode Split 66%

Selanjutnya Hasil perbandingan kedua metode uji yang diperoleh berdasarkan output pada tool WEKA ditunjukkan pada Tabel 6 berikut ini.

Tabel 6. 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.

3 Simpulan

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 menggunakan dataset intruksi 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 0,0122 % untuk klasifikasi serangan dengan metode Split 66 % dan 10 Fold Cross Validation dengan waktu 1,03 detik dan 1,08 detik.

DAFTAR PUSTAKA

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

[2] 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.

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

[4] Symantec Corporation, "Internet Security Threat Report 2019 Appendices," Internet Secur. Threat Rep., vol. 24, February, 2019.