Jurnal Teknologi Sistem Informasi dan Aplikasi **ISSN: 2654-3788**
Vol. 2, No. 4, Oktober 2019 (135-140) **e-ISSN: 2654-4229**

Klasifikasi Anomali Intrusion Detection System (IDS) Menggunakan Algoritma Naïve Bayes Classifier dan Correlation-Based Feature Selection

Saipul Anwar[1], Fajar Septian[2], Ristasari Dwi Septiana[3]

[1]Sistem Informasi, Universitas Tanri Abeng, Jakarta Selatan, Indonesia
[2]Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Banten, Indonesia
[3]Teknik Informatika, STMIK Eresha, Tangerang Selatan, Banten, Indonesia
e-mail: [1]saipul@tau.ac.id, [2]dosen00677@unpam.ac.id, [3]ristasari.dwis@gmail.com

Abstract

Intrusion Detection System (IDS) is useful for detecting an attack or disturbance on a network or information system. Anomaly detection is a type of IDS that can detect a deviate attack on the network based on statistical probability. The increasing use of the internet also increases interference or attacks from intruders or crackers that exploit weak internet protocols and application software. When many data packets arrive, a problem arises that needs to be analyzed. The right technique to analyze the data package is data mining. This study aims to classify IDS anomalies using the Naïve Bayes classification algorithm from the results of attribute selection with correlation-based feature selection. This study uses a UNSW-NB15 intrusion detection system data collection consisting of 49 attributes and 321,283 data records. Performance measurements are based on accuracy, precision, F-Measure and ROC Area. The results of attribute selection with correlation-based feature selection leave 4 attributes. The results of the evaluation of IDS anomaly classification using the naïve Bayes algorithm without the precedence of the attributes selected by the correlation technique obtained an accuracy rate of 71.2%. While the classification results if preceded by the attributes selected by the correlation technique obtained an accuracy of 74.8%. Classification with the naïve Bayes algorithm can be improved its accuracy which is preceded by the selection of attributes with correlation techniques.

Keywords: correlation-based fetaure selection, classification, data mining, intrusion detection system, naïve bayes

Abstrak

Intrusion Detection System (IDS) berguna untuk mendeteksi adanya serangan atau gangguan pada suatu jaringan atau sistem informasi. Deteksi anomali merupakan salah satu jenis IDS yang mendeteksi serangan menyimpang pada jaringan berdasarkan probabilitas statistik. Meningkatnya penggunaan internet maka meningkat juga gangguan atau serangan dari *intruder* atau *cracker* yang mengeksploitasi protokol internet dan software aplikasi yang lemah. Pada saat banyaknya paket data yang datang, muncul masalah yang perlu dilakukan analisa. Teknik yang tepat untuk menganalisa paket data tersebut adalah *data mining*. Penelitian ini bertujuan untuk mengklasifikasikan anomali IDS menggunakan algoritma klasifikasi Naïve Bayes dari hasil pemilihan atribut dengan teknik korelasi (*correlation-based feature selection*). Penelitian ini menggunakan koleksi data *intrusion detection system* UNSW-NB15 yang terdiri dari 49 atribut dan 321.283 *record* data. Pengukuran performa didasarkan pada akurasi, presisi, *F-Measure* dan *ROC Area*. Hasil seleksi atribut dengan *correlation-based feature selection* meninggalkan 4 atribut. Hasil evaluasi klasifikasi anomali IDS menggunakan algoritma naïve bayes tanpa didahului atribut yang diseleksi dengan teknik korelasi diperoleh tingkat akurasi 71,2 %. Sedangkan hasil klasifikasi jika didahului dengan atribut yang diseleksi dengan teknik korelasi didapatkan akurasi 74,8 %. Klasifikasi dengan algoritma naïve bayes dapat ditingkatkan akurasinya yang didahului pemilihan atribut dengan teknik korelasi.

Kata kunci: *correlation-based fetaure selection, data mining, intrusion detection system,* klasifikasi, naïve bayes

1. Pendahuluan

Salah satu aspek penting dalam internet khususnya jaringan komputer adalah keamanan. Suatu jaringan komputer harus dapat memberikan keamanan kepada *user* terhadap akses yang dilakukannya, dan memberikan jaminan bahwa informasi dan/atau data pribadinya aman dari *intruder* (penyerang). Internet yang terus berkembang secara eksponensial berakibat juga pada meningkatnya gangguan atau serangan *intruder* atau *cracker* untuk mengeksploitasi protokol *internet* dan aplikasi yang rentan terhadap serangan. Jaringan komputer akan terus mengalami peningkatan serangan, terutama melalui internet dari tahun ke tahun. Informasi yang diberikan oleh *Kaspersky Lab* pada tahun 2007 terdapat 23.680.646 seranga melalui *browser internet*, kemudian terjadi peningkatan serangan pada tahun 2009 ke angka 73.619.767 serangan dan terus meningkat ke angka 580.371.937 serangan tahun 2010. Menyebarnya program-program *malicious* di antara para pengguna komputer pada tahun 2010 melalui *internet browser*. Hanya 60% serangan *web* yang dapat dideteksi oleh *Kaspersky Security Network* (Gostev & Namestnikov, 2011).

Intrusion detection pada sistem komputer atau jaringan dilakukan dengan cara memantau adanya anomali yang terjadi, selanjutnya dilakukan analisa dan mengeluarkan peringatan adanya pelanggaran atau mendekati pelanggaran terhadap kebijakan keamanan komputer atau praktik keamanan standar (Scarfone, K, 2007). *Intrusion detection system* (IDS) digunakan untuk mengidentifikasi lalu lintas paket-paket data yang ditransmisikan melalui jaringan computer, selanjutnya menentukan paket-paket data tersebut aman, mencurigakan atau merupakan sebuah serangan.

Masalah muncul ketika terdapat aktifitas-aktifitas mencurigakan atau aktifitas tersebut adalah serangan tetapi tidak terdaftar pada aturan keamanan yang terdaftar, sehingga hal tersebut sangat berbahaya bagi jaringan komputer. Oleh karena itu dibutuhkan sebuah sistem klasifikasi serangan yang berfungsi untuk mengklasifikasi anomali lalu lintas jaringan yang ada dan dari klasifikasi tersebut akan diketahui apakah sebuah aktifitas pada jaringan tersebut adalah serangan atau bukan serangan. Dari hasil klasifikasi tersebut juga dapat digunakan menjadi dasar untuk membuat aturan baru yang akan didaftarkan pada aplikasi IDS yang digunakan.

2. Tinjauan Pustaka

Dalam penelitian ini meninjau beberapa penelitan terkait sebagai data pendukung dan pembanding hasil penelitan. Peneliti juga merujuk beberapa pustaka sebagai landasan dilakukannya penelitian ini.

2.1 Penelitian Terkait

Untuk mendapatkan data yang lebih lengkap, peneliti merujuk beberapa penelitian terkait yang pernah dilakukan peneliti lain. Hasil penelitian tersebut juga digunakan sebagai komparasi dengan penelitian yang dilakukan agar mendapat hasil yang lebih baik.

Penelitian *data minig* pada bidang pendidikan (*Educational Data Mining*) tentang prediksi kinerja mahasiswa dengan komparasi model klasifikasi *Naive Bayes* dan *Decision Tree* C4.5 (Galih, 2019). *Data set* yang digunakan adalah *dataset* akademik mahasiswa STMIK Jabar dengan delapan atribut. Hasil pengujian pada model *Naive Bayes didapatkan* akurasi 86,83% dengan *ratio data training* 80% dan model *Decision Tree* C4.5 88,10% dengan *ratio data training* 90%.

Penelitian mengklasifikasi serangan terhadap *intrusion detection system* pada bidang jaringan menggunakan algoritma *Decision Tree* C4.5 (Khaerani & Handoko, 2015). *Dataset* yang digunakan adalah KDD'99 yang memiliki 41 atribut. Atribut diseleksi menggunakan teknik evolusi dan didapatkan 16 atribut yang relevan. Dengan *feature selection evoluationary* penelitian ini menghasilkan nilai akurasi sebesar 98.67%.

Penelitian *data mining* pada bidang jaringan dengan menerapkan algoritma *Naive Bayes Classifier* (NBC) diskritisasi variabel (Wirawan & Eksistyanto, 2015). *Dataset* yang digunakan adalah NSL-KDD99. Teknik rata-rata atau simpangan baku dipakai untuk seleksi atribut kontinyu berdasar korelasi dengan diskritisasi 3 interval dan 5 interval. Dengan diskritisasi dapat meningkatkan akurasi mencapai 89% dan rata-rata *running time* proses klasifikasi 31 detik.

2.2 Data Mining

Kajian yang membahas tentang teknik-teknik mendapatkan pengetahuan berdasarkan hasil penemuan pola-pola tertentu dari kumpulan data yang berukuran besar (Han, Kamber, & Pei, 2012). *Knowledge Discovery In Database* (KDD) merupakan istilah lain dari *data mining*, yaitu aktifitas mengumpulkan dan memakai data yang berukuran besar dalam rangka mendapatkan pola atau hubungan data. Keluaran *data mining*,

selanjutnya dapat digunakan untuk pengambilan keputusan di waktu mendatang (Santosa, 2007).

Urutan proses KDD adalah sebagai berikut (Han, Kamber, & Pei, 2012) :

1. Pembersihan Data (*Data Cleaning*)
   Dilakukan untuk membuang *noise* dan data yang inkonsisten. Penghapusan data yang kelengkapan atributnya tidak sesuai dengan yang dibutuhkan.
2. Integrasi Data (*Data Integration*)
   Dilakukan untuk mengkombinasikan beberapa sumber data. Data dari berbagai sumber digabungkan menjadi satu tempat penyimpanan data.
3. Seleksi Data (*Data Selection*)
   Mengambil atau memilih data yang memiliki keterkaitan dengan teknik analisis basis data. Meminimalkan representasi data dan informasi data yang hilang menggunakan teknik atau metode tertentu.
4. Transformasi Data (*Data Transformation*)
   Mengubah bentuk data dengan meringkas atau menggabungkan data agar sesuai untuk tahap *mining*.
5. Penambangan Data (*Data Mining*)
   Mendapatkan pola atau informasi yang menarik dari data yang sudah diseleksi berdasarkan tujuan awal menggunakan algoritma tertentu.
6. Evaluasi Pola (*Pattern Evaluation*)
   Mengidentifikasi pola yang didapatkan yang akan dijadikan sebagai pengetahuan.
7. Representasi Pengetahuan (*Knowledge Presentation*)
   Pola atau pengetahuan yang didapatkan selanjutnya disajikan secara visual agar dapat dipahami.

2.3 Intrusion Detection System (IDS)

Intrusion Detection System (IDS) merupakan bentuk perlakuan terhadap lalu lintas paket data pada sebuah jaringan atau *device* untuk mendeteksi adanya tindakan yang dianggap mencurigakan atau tidak diinginkan. IDS memungkinkan untuk diterapkan ke dalam sebuah *device* dengan aplikasi tertentu, dan selanjutnya aplikasi tersebut akan mendeteksi paket data *illegal* yang melalui suatu jaringan seperti paket yang berbahaya terhadap kebijakan keamanan dan menerobos atau mencuri autentikasi pengguna (Wu, 2009).

Deteksi intrusi dilakukan dengan mengamati dan menganalisa semua peristiwa pada sebuah jarigan komputer terhadap gejala-gejala terjadinya permasalahan keamanan. Pada umumnya, IDS akan memberikan peringatan kepada *administrator* sistem apabila terdeteksi terjadi intrusi. Tantangan IDS adalah (Lazarević, Srivastava, & Kumar, 2018):

1. Besarnya ukuran data.
2. Tingginya dimensi data.
3. Hubungan kedekatan waktu data.
4. Tidak seimbangnya distribusi kelas
5. Analisis *preprocessing* data yang terkumpul.
6. Kemampuan komputasi yang tinggi, terutama IDS daring, skalabilitas dan terdistribusi.

2.4 Klasifikasi

Klasifikasi adalah metode yang dipakai untuk mencari sekelompok *model* (fungsi) sebagai deskripsi dan pembeda antar kelas-kelas data agar *model* tersebut dapat digunakan untuk memprediksi objek yang belum diketahui kelasnya atau memprediksi kecondongan data-data yang dihasilkan di waktu mendatang (Han, Kamber, & Pei, 2012). Klasifikasi mempunyai dua tugas utama, yaitu membangun *model* sebagai contoh, dan melakukan identifikasi/prediksi berdasarkan model yang sudah dibuat terhadap objek data baru yang dihasilkan di waktu mendatang, berada pada kelas mana kah objek data baru tersebut.

2.5 Naive Bayes Classifier (NBC)

Naive bayes Classifier selanjutnya disebut NBC termasuk teknik prediksi berdasarkan probabilitas sederhana pada teorema Bayes (Prasetyo, 2012). *Naive bayes* adalah teknik penalaran probabilitas melalui kumpulan probabilitas yang dihitung dan menjumlahkan frekuensi dan kombinasi koleksi data (Galih, 2019). Nilai probabilitas dalam metode ini digunakan sebagai penentuan keputusan karena setiap kasus terdapat proses komputasi resiko. Persamaan *Naïve Bayes* diperoleh dari rumus *bayes* berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

X = Data di mana kelasnya belum teridentifikasi
H = Hipotesis di mana kelasnya sudah teridentifikasi
P(H|X) = Probabilitas H dalam X (persentase banyak H dalam X)
P(X|H) = Probabilitas X dalam H (persentase banyak X dalam H)
P(H) = Probabilitas *Prior* H
P(X) = Probabilitas *Prior* X

2.6 Correlation-Based Feature Selection

Tahap pemilihan atribut dalam koleksi data yang memiliki keterkaitan dengan kelasnya (Wirawan & Eksistyanto, 2015). Nilai korelasi dan *mutual information* adalah fitur yang biasa digunakan dalam pemilihan atribut. Nilai korelasi dan *mutual information* termasuk fitur yang juga diterapkan pada teknik *Corellation-based Feature Selection* (CFS). Pemilihan atribut dilakukan melalui perhitungan keterkaitan atribut dengan kelasnya serta keterkaitan atribut tertentu dengan atribut yang lain. Atribut yang terpilih adalah nilai keterkaitan atribut dengan kelasnya tinggi, tetapi keterkaitan atribut dengan atribut yang lain rendah. CFS dinotasikan sebagai berikut:

$$r_{zc} = \frac{kr_{zi}}{\sqrt{k+k(k-1)rii}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

r_{zc} = hubungan antara banyaknya atribut dengan variabel yang berada di luar kelas
k = Banyak atribut
r_{zi} = rata-rata hubungan antara atribut dengan variabel yang berada di luar kelas
rii = rata-rata hubungan antar atribut

3. Metode Penelitian

Penelitian yang dilakukan ialah menerapkan algoritma klasifikasi *naïve bayes* yang didahului dengan pemilihan atribut menggunakan teknik korelasi atribut untuk mendeteksi anomali sistem pendeteksi gangguan atau serangan pada jaringan computer dan menyajikan hasil kerja algoritma tersebut. Hasil kerja algoritma diukur dengan melihat nilai akurasi, presisi, *f-measure* dan *ROC area*. Koleksi data yang digunakan pada penelitian ini dibagi menjadi dua, yaitu data pelatihan dan data pengujian. 10-*Fold Cross-Validation* digunakan untuk evaluasi pengujian, di mana data selanjutnya dibagi lagi ke dalam sepuluh bagian. Sembilan bagian digunakan sebagai data pelatihan dan satu sebagai data pengujian, dan seterusnya hingga tiap-tiap bagian dipakai sebagai data pengujian. Nilai akursi pengujian yang diperoleh ialah rata-rata dari nilai akurasi tiap-tiap bagian.

Koleksi data yang dipakai pada penelitian ini ialah UNSW-NB15 tahun 2015. UNSW-NB15 merepresentasikan sembilan besar mayoritas serangan dengan menggunakan *IXIA PerfectStorm Tool* dari simulasi yang dilakukan dengan periode waktu 16 jam pada 22 Januari 2015 dan 15 jam pada 17 Februari 2015 untuk merekam 100 GBs data. Ada 49 atribut yang telah dihasilkan dengan menggunakan Argus, Bro-IDS tool dan dua belas algoritma yang dibangun dengan bahasa C# yang mencakup karakteristik paket jaringan (Moustafa & Slay, 2015). Dari dataset awal sebanyak 2.540.044 *record* diambil sampling sebanyak 321.283 *record* data. Dari data tersebut terdapat 49 atribut sebagai berikut:

Tabel 1 Atribut Dataset UNSW-NB15

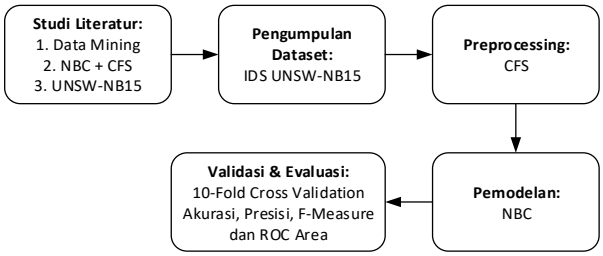
No.	Name	Type	25	trans_depth	I
1	srcip	N	26	res_bdy_len	I
2	sport	I	27	sjit	F
3	dstip	N	28	djit	F
4	dsport	I	29	stime	T
5	proto	N	30	ltime	T
6	state	N	31	sintpkt	F
7	dur	F	32	dintpkt	F
8	sbytes	I	33	tcprtt	F
9	dbytes	I	34	synack	F
10	sttl	I	35	ackdat	F
11	dttl	I	36	is_sm_ips_ports	B
12	sloss	I	37	ct_state_ttl	I
13	dloss	I	38	ct_flw_http_mthd	I
14	service	N	39	is_ftp_login	B
15	sload	F	40	ct_ftp_cmd	I
16	dload	F	41	ct_srv_src	I
17	spkts	I	42	ct_srv_dst	I
18	dpkts	I	43	ct_dst_ltm	I
19	swin	I	44	ct_src_ ltm	I
20	dwin	I	45	ct_src_dport_ltm	I
21	stcpb	I	46	ct_dst_sport_ltm	I
22	dtcpb	I	47	ct_dst_src_ltm	I
23	smeansz	I	48	attack_cat	N
24	dmeansz	I	49	label	B

Koleksi data yang diapakai hanya data dengan label anomali atau tidak normal yang seluruhnya berjumlah 321.283 *records*. Sedangkan data normal sebanyak 2.218.761 *records* tidak digunakan. Berikut adalah distribusi data yang digunakan:

Tabel 2 Distribusi Dataset UNSW-NB15

Type	*Records*
Fuzzers	*24,246*
Analysis	*2,677*
Backdoors	*2,329*
DoS	*16,353*
Exploits	*44,525*
Generic	*215,481*
Reconnaissance	*13,987*
Shellcode	*1,511*
Worms	*174*

Penulis menyusun desain penelitian berupa tahapan-tahapan penelitian agar penelitian dapat dilakukan secara sistematis. Berikut adalah gambar desain penelitian ini:



Gambar 2 Desain Penelitian

4. Hasil Penelitian

Dataset awal yang mememiliki 49 atribut kemudian dengan *Correlation-based Feature Subset* (CFS) *Selection*, dengan *search method Greedy Stepwise* dan dengan *Attribut Selection Mode* 10-*Fold Cross Validation* untuk mengurangi atribut yang dianggap tidak perlu menggunakan aplikasi WEKA meninggalkan 4 atribut yaitu *sport* (100%) , *dsport* (100%), *sbytes* (100%), dan *service* (100%).

```
=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

number of folds (%)  attribute
          0(  0 %)     1 srcip
         10(100 %)     2 sport
          0(  0 %)     3 dstip
         10(100 %)     4 dsport
          0(  0 %)     5 proto
          0(  0 %)     6 state
          0(  0 %)     7 dur
         10(100 %)     8 sbytes
          0(  0 %)     9 dbytes
          0(  0 %)    10 sttl
          0(  0 %)    11 dttl
          0(  0 %)    12 sloss
          0(  0 %)    13 dloss
         10(100 %)    14 service
          0(  0 %)    15 Sload
          0(  0 %)    16 Dload
          0(  0 %)    17 Spkts
          0(  0 %)    18 Dpkts
          0(  0 %)    19 swin
          0(  0 %)    20 dwin
          0(  0 %)    21 stcpb
          0(  0 %)    22 dtcpb
          0(  0 %)    23 smeansz
          0(  0 %)    24 dmeansz
          0(  0 %)    25 trans_depth
          0(  0 %)    26 res_bdy_len
          0(  0 %)    27 Sjit
          0(  0 %)    28 Djit
          0(  0 %)    29 Stime
          0(  0 %)    30 Ltime
          0(  0 %)    31 Sintpkt
          0(  0 %)    32 Dintpkt
          0(  0 %)    33 tcprtt
          0(  0 %)    34 synack
          0(  0 %)    35 ackdat
          0(  0 %)    36 is_sm_ips_ports
          0(  0 %)    37 ct_state_ttl
          0(  0 %)    38 ct_flw_http_mthd
          0(  0 %)    39 is_ftp_login
          0(  0 %)    40 ct_ftp_cmd
```

Gambar 3 Hasil Seleksi Atribut dengan CFS

Penjelasan setiap atribut hasil seleksi menggunakan teknik *Correlation-based Feature Subset (CFS) Selection* disajikan pada table 3.

Tabel 3 Hasil Seleksi Atribut dengan CFS

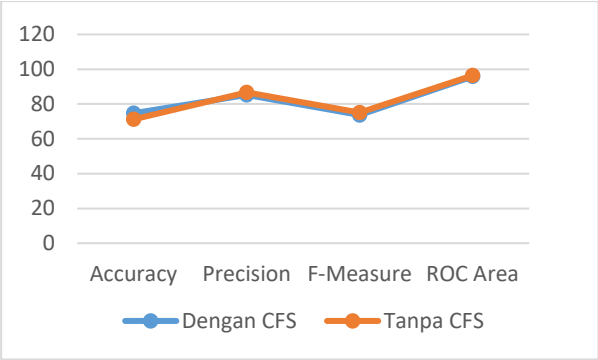
Nama	Tipe	Deskripsi
sport	I	*port* sumber
dsport	I	*port* tujuan
sbytes	I	Besar transaksi data dari sumber ke tujuan (*bytes*)
service	N	http, ftp, smtp, ssh, dns, ftp-data, irc dan (-) jika layanan tidak banyak digunakan

Pemodelan data menggunakan algoritma klasifikasi naïve bayes untuk melakukan klasifikasi terhadap IDS. Terdapat lima atribut yang digunakan untuk pemodelan data, yaitu *sport*, *dsport*, *sbytes*, *service*, dan *attack_cat*. Hasil pemodelan data menggunakan aplikasi WEKA tanpa menggunakan seleksi atribut membutuhkan waktu proses 2,78 detik. Sedangkan waktu proses pemodelan data dengan menggunakan seleksi atribut *Correlation-based Feature Selection* (CFS) adalah 0,45 detik, lebih cepat 6 kali lipat dari pemodelan tanpa seleksi atribut.

Tabel 4 Hasil Pengukuran Algoritma Naïve Bayes

	Accuracy	*Precision*	*F-Measure*	*ROC Area*
Dengan CFS	*74,8*	*85,2*	*73,7*	*95,8*
Tanpa CFS	*71,2*	*86,7*	*75,1*	*96,4*

Hasil eksperimen yang ditunjukan tabel 4, akurasi kinerja dari algoritma *naïve bayes* dengan *Correlation-based Feature Selection* (CFS) lebih baik dari pada tanpa *Correlation-based Feature Selection* (CFS). Nilai akurasi *naïve bayes* tanpa CFS sebesar 71,2% sedangkan dengan CFS sebesar 74,8%. Nilai presisi *naïve bayes* tanpa CFS sebesar 86,7% sedangkan dengan CFS sebesar 85,2%. Nilai *F-Measure naïve bayes* tanpa CFS sebesar 75,1% dan dengan CFS sebesar 73,7%. Nilai *ROC Area naïve bayes* tanpa CFS sebesar 96,4% dan dengan CFS sebesar 95,8%. Dari empat kriteria pengukuran hanya nilai akurasi yang dapat memberikan hasil yang lebih baik. Dari hasil pengujian terlihat bahwa kinerja klasifikasi yang dilakukan algoritma *naïve bayes* dengan CFS lebih baik dari pada tanpa CFS. Atribut yang dipilih dengan teknik CFS berdampak pada peningkatan kinerja algoritma *naïve bayes* ketika melakukan klasifikasi anomali IDS pada koleksi data UNSW-NB15.



Gambar 4 Grafik Perbandingan Hasil Pengukuran

5. Kesimpulan

Berdasarkan hasil pengujian algoritma *naïve bayes* untuk klasifikasi anomali IDS yang diawali pemilihan atribut dengan teknik korelasi diperoleh kesimpulan sebagai berikut:

a. Pemodelan data menggunakan aplikasi WEKA tanpa menggunakan CFS membutuhkan waktu proses 2,78 detik, lebih cepat 6 kali lipat lebih cepat dari pemodelan tanpa menggunakan CFS 0,45 detik.
b. Nilai akurasi *naïve bayes* tanpa *Correlation-based Feature Selection* (CFS) sebesar 71,2%, lebih kecil dari pada akurasi *naïve bayes* dengan *Correlation-based Feature Selection* (CFS), yaitu sebesar 74,8%.
c. Pemilihan atribut dengan teknik *Correlation-based Feature Selection* (CFS) berdampak pada peningkatan akurasi algoritma *naïve bayes* pada koleksi data *Intrusion Detection System* UNSW-NB15.

Daftar Pustaka

Galih. (2019). Data Mining di Bidang Pendidikan untuk Analisa Prediksi Kinerja Mahasiswa dengan Komparasi 2 Model Klasifikasi pada STMIK Jabar. *Jurnal Teknologi Sistem Informasi dan Aplikasi*, *2*(1), 23-30.

Gostev, A., & Namestnikov, Y. (2011, Februari 17). *Kaspersky Security Bulletin 2010. Statistics, 2010.* Retrieved Agustus 10, 2019, from Kaspersky Securelist: https://securelist.com/kaspersky-security-bulletin-2010-statistics-2010/36345/

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques Third Edition.* USA: Elsevier.

Khaerani, I., & Handoko, B. (2015). Implementasi Dan Analisa Hasil Data Mining Untuk Klasifikasi Serangan Pada Intrusion Detection System (IDS) Dengan Algoritma C4.5. *Techno.COM*, *14*(3), 181-188.

Lazarević, A., Srivastava, J., & Kumar, V. (2018, August 3). *Data Mining For Intrusion Detection Tutorial on the Pacific-Asia Conference on Knowledge Discovery in Databases 2003.* Retrieved August 10, 2019, from iDoc Slide: https://idocslide.org/document/data-mining-for-intrusion-detection-tutorial-on-the-pacific-asia-conference-on-knowledge-discovery-in-databases-2003

Moustafa, N., & Slay, J. (2015). UNSW-NB15: A Comprehensive Data set for Network Intrusion Detection systems (UNSW-NB15 Network Data Set). *Military Communications and Information Systems Conference (MilCIS).* IEEE.

Prasetyo, E. (2012). *Klasifikasi Naive bayes.* Jawa Timur: Teknik Informatika, Universitas Pembangunan Nasional "Veteran".

Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis.* Yogyakarta: Graha Ilmu.

Wirawan, I. T., & Eksistyanto, I. (2015). Penerapan Naive Bayes Pada Intrusion Detection System Dengan Diskritisasi Variabel. *JUTI: Jurnal Ilmiah Teknologi Informasi*, *13*(2), 182-189.

Wu, T. (2009). *Information Assurance Tools Report – Intrusion Detection Systems Sixth Edition.* Defense Technical Information Center, Information Assurance Technology Analysis Center (IATAC). Herndon, United States: IATAC.