DETEKSI SERANGAN PADA INTRUSION DETECTION SYSTEM (IDS) UNTUK KLASIFIKASI SERANGAN DENGAN ALGORITMA NAÏVE BAYES, C.45 DAN K-NN DALAM MEMINIMALISASI RESIKO TERHADAP PENGGUNA

Niko Suwaryo[1], Ismasari Nawangsih[2], Sri Rejeki[3]
[1]Universitas Medika Suherman, [2]Universitas Pelita Bangsa, [3]Universitas Bhayangkara Jakarta Raya
[1]suwaryoniko@gmail.com, [2]ismasari.n@pelitabangsa.ac.id,
[3]sri.rejeki@dsn.ubharajaya.ac.id

ABSTRACT

Intrusion Detection System is the ability possessed by hardware or software that serves to detect suspicious activity on the network and analyze and search in general. The purpose of this study is to classify attack detection on the Intrusion Detection System using the C.45, Naïve Bayes and K-NN algorithms to see how big the attack is. The benefits gained in this study are as a test and learning material in analyzing, classifying attacks so that they can prevent and minimize attacks to users. To overcome this problem, this study uses the C.45 algorithm, Naïve Bayes, K-NN, K-NN algorithm produces an accuracy rate of 82.58%, Recall 81.73% and Precision 84.11% while the Naïve Bayes accuracy 96.91%, Recall 97,45% and Percision 96.18% and the algorithm produces an optimal value of C.45 accuracy 97.80% Recall 98.18% and Precision 97.60%. On the attribute (attack) which has the number of classes or normal labels, dos, probes, r21. The results of the lowest K-NN algorithm are caused or normal to be considered yes(an attack) which should be No(no attack)and the C.45 algorithm attribute(attack) normal, dos, probe and r21, normal(no attack), yes(the presence of an attack) is optimal in the classification of attack detection data on Intrusion Detection System(IDS).

Keywords: Data Mining, C.45, Naïve Bayes and K-NN, Intrusion Detection System(IDS)

I. PENDAHULUAN

Data Pada komputer jaringan dapat diproses ke dalam komputer, serangan komputer jaringan upaya mendapatkan akses pada komputer, *Intrusion Detection System* dapat digunakan dalam membaca aktivitas yang dicurigai dari suatu sistem komputer jaringan dan mendeteksi aktivitas komputer jaringan yang dapat masuk ke dalam suatu komputer jaringan dalam mencari celah atau kelemahan dari suatu komputer untuk melakukan serangan atau percoban yang akan dianalisa apakah ada penyusupan atau percobaan serangan, komputer jaringan yang terletak pada jaringan segmen server atau berada pada pintu masuk pada suatu komputer jaringan. *Intrusion Detection System* merupakan fungsi yang dimiliki atau kemampuan dari perangkat keras untuk mencari dan menganalisa, mendeteksi suatu aktivitas yang dicurigai pada komputer jaringan pada umumnya, *Intrusion Detection System* ada tiga bentuk yang saat ini masih dipakai ketiganya masih mempunyai perbedaan dalam mendeteksi dan mencegah aktivitas jahat, ketiganya dapat dikembangkan dalam menghasilkan hasil yang optimal dan efektif dalam mencegah serangan atau penyusupan dalam suatu

komputer jaringan. (Kharisma Muchammad, 2017)

Intrusion Detection System (IDS) merupakan proses memonitor trafik jaringan dalam sebuah sistem untuk mendeteksi adanya pola dan aktivitas yang mencurigakan yang memungkinkan adanya serangan dalam suatu sistem tersebut. Terdapat dua kategori teknik yang digunakan untuk deteksi *intrusi* yaitu deteksi *intrusi* berbasis *anomaly* dan deteksi *intrusi* berbasis, perkembangan teknologi komputer semakin pesat sehingga mengalami perubahan dibidang teknologi . Salah satu manfaat komputer sendiri untuk mempermudah dalam pengoprasian data dapat ditampilkan dalam bentuk informasi dan informasi ini disimpan dalam satu komputer dan keamanan jaringan komputer salah satu faktor penting dalam dunia teknologi informasi dan sebagai perlindungan informasi dan data-data yang dianggap penting oleh suatu lembaga atau perusahaan.

Data dimanfaatkan untuk penelitian dan dapat diolah menjadi sebuah pengetahuan, informasi dan data serangan pada *Intrusion Detection System* dapat diklasifikasi, mining data merupakan semi otomatik dalam menerapkan matematika, dan kecerdasan buatan untuk menguraikan atau mengetahui informasi dengan teknik yang diharapkan dan dapat digali suatu potensi informasi atau pengetahuan serta dapat menganalisa serangan terhadap keamanan dengan *Intrusion Detection System (IDS)* pada komputer tersebut, untuk meningkatkan pengetahuan dan selain itu bisa digunakan sebagai sarana untuk mengambil keputusan dalam meningkatkan informasi terhadap pengguna sistem komputer yang digunakan.

Berdasarkan uraian masalah diatas maka dalam penelitian ini mengambil data mining dengan *algoritma K-NN, Naïve Bay*es dan *C.45* untuk mengetahui seberapa besar *intrusi* pada serangan deteksi sistem. Algoritma *K-NN, Naïve Bayes dan C.45* salah suatu algoritma yang dapat digunakan dalam *classifier* suatu dataset yang menjadi suatu pengetahuan dan informasi, teknik dari *classifier* bagian dari sebuah algoritma yang dapat diuji atau bagian dari data mining yang dapat diprediksi dari sebuah data *Intrusion Detection System (IDS)* berdasarkan sekumpulan data dari atribut-atribut.

Intrusion Detection system (IDS)

Intrusion Detection System (IDS) adalah proses pemantauan peristiwa yang terjadi pada sistem komputer atau jaringan dan menganalisisnya untuk menentukan kegiatan ini, termasuk normal atau intrusi. Model proses IDS memiliki 3 fungsi dasar, yaitu: pertama, pengambilan data dari berbagai tingkatan sistem seperti jaringan, host, dan aplikasi. Intrusion Detection system (IDS) adalah kemampuan yang dimiliki oleh perangkat keras atau perangkat lunak yang berfungsi untuk mendeteksi aktivitas mencurigakan pada jaringan dan menganalisis dan mencari Secara umum, IDS dibagi menjadi dua bentuk yang digunakan saat ini dan keduanya me-miliki perbedaan dalam hal mendeteksi dan menangguhkan kejahatan. kegiatan. Keduanya harus dikembangkan, sehingga hasilnya lebih efektif dalam mendeteksi setiap infiltrasi dan menyiapkan strategi yang tepat. Berikut adalah tiga bentuk Sistem Deteksi Intrusi (IDS) (Kharisma Muchammad,2017).

1. Network-Intrusion Detection System (NIDS) Berbasis Jaringan adalah jaringan komputer yang dapat dilihat oleh komputer atau jaringan, bagian ini dapat digunakan untuk lebih

efektif dalam mengetahui grafik keluar atau masuk, dihosting dalam grafik atau segmen antara lokal jaringan. Sistem deteksi serangan Berbasis-Jaringan dapat dikembangkan di belakang dan di depan VPN getways atau firewall untuk mengukur efektivitas perangkat lunak dalam keamanan dan berinteraksi untuk memperkuat keamanan jaringan.

2. Sistem deteksi serangan Berbasis-Host Intrusion adalah aplikasi perangkat lunak khusus yang dapat diinstal pada komputer (server) untuk dilihat dalam semua komunikasi keluar atau masuk dari server dan dalam memonitor sistem data jika ada perubahan.
3. Distributed Intrusion Detection System (IDS) adalah sensor yang terhubung satu sama lain dapat berfungsi untuk sensor jarak jauh untuk menyediakan pelaporan ke sistem pusat.

Dataset KDD99

KDD adalah kumpulan set data atau data yang dapat digunakan dalam menganalisis data untuk memberikan informasi atau pengetahuan tentang data, protokol transfer contorol dari suatu sistem pada jaringan dalam mengetahui aktivitas jaringan dan serangan terhadap suatu sistem atau normal, pengambilan data masih dilakukan secara manual belum otomatis, dataset besar dalam 4GB dalam format zip (Kharisma Muchammad,2017)

1. Root to User (U2R): label serangan pada sistem dapat diuraikan secara umum untuk mengakses untuk mendapatkan admin di komputer, serangan mencari kelemahan terbuka untuk mendapatkan akses untuk menyerang atau menyadaf suatu data, file yang ada di komputer.
2. Remote to Local (R2L): serangan yang ingin selalu mencari cacat atau di kontrol komputer yang akan digunakan untuk mengambil data.
3. Probe atau probing: kelas serangan ini umumnya bertujuan untuk mencari informasi tentang jaringan komputer yang akan diserang. Ping untuk memeriksa apakah komputer dengan IP tertentu ada atau tidak, contoh lain adalah serangan pemindaian port untuk melihat port mana yang terbuka dari komputer atau dalam dataset pelatihan. Probabilitas distribusi serangan berbeda dari probabilitas kemunculan dalam data pelatihan. beberapa ahli melihat bahwa serangan baru secara umum adalah variasi serangan yang telah diidentifikasi sebelumnya. Sehingga penggunaan tanda tangan lama serangan IDS masih bisa digunakan untuk mendeteksi variasi serangan baru.

Algoritma Naïve Bayes

Algoritma Naive Bayes Pengklasifikasi bayesian adalah pengklasifikasi statistik dan didasarkan pada teorema bayes. Teori keputusan bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (pattern recognition), penggunaan algoritma ini dalam hal klasifikasi harus mempunyai masalah yang bisa dilihat statistiknya. Misalkan X adalah set atribut data dan h kelas variabel dan jika kelas memiliki hubungan dengan atribut maka diperlukan X dan h sebagai variabel acak dan menangkap hubungan peluang P(h|X) ini peluang posterior untuk h dan sebaliknya perior P(h) ((Suryanto,2017).

Naive Bayes mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah independen secara bersyarat yang diberikan dengan label kelas labelkelas label kelas y dengan tiap

set atribut X = {X1 , X 2 ,2, X d } terdiri dari d atribut. Tahapan algoritma naive bayes:

1. Menyiapkan data training.
2. Setiap data dipresentasikan sebagai vektor berdimensi-n yaitu $X = X_1, X_2, X_3 \ldots . Xn$
3. N adalah gambaran dari ukuran yang dibuat di test dari n atribut yaitu $A_1, A_2, A_3 \ldots . An$
4. M adalah kumpulan kategori yaitu $X = C_1, C_2, C_3 \ldots . Cm$
5. Diberikan data test X yang tidak diketahui kategorinya, maka classifier akan memprediksi bahwa X adalah milik kategori dengan posterior probability tertinggi berdasarkan kondisi X.
6. Naive bayes classifier menandai bahwa test X yang tidak diketahui tadi ke kategori C1 jika dan hanya jika P(Ci|X)>P(Cj|X) untuk $1 \leq j \leq m, j \neq i$
7. Kemudian kita perlu memaksimalkan P(Ci|X) P(Ci|X)= P(X|Ci). P(Ci) P(X).
8. Dimana x adalah nilai-nilai atribut dalam sampel X dan probabilitas P(x1|Ci), P(x2|Ci),..... P(xn|Ci), dapat diperkirakan dari data training

$$P(C_i \mid X) = \frac{P(X \mid C_i) P(C_i)}{P(X)}$$

$$P(C_i \mid X) = P(X \mid C_i) P(C_i)$$

$$P(X \mid C_i) = \prod_{k=1}^{n} P(x_k \mid C_i) = P(x_1 \mid C_i) \times P(x_2 \mid C_i) \times \ldots \times P(x_n \mid C_i)$$

Decision Tree (C.45)

Decision Tree adalah salah stau metode klasifikasi yang populer dan banyak digunakan secara praktis. Metode ini berusaha menemukan model klasifikasi yang tahan terhadap derau, salah satu metode decision tree yang sangat populer adalah *iteractive dychotomizer version* 3 (ID3) dua varian yang lainnya sangat populer adalah C.45 dan Assistant (Suryanto, 2017).

Berikut ini tahapan proses permodelan dalam penelitian ini. Metode prediksi, Algoritma C.45 dipilih karena salah satu kelebihannya adalah dapat menangani data numerik dan diskret. Algoritma C.45 menggunakan rasio perolehan (*gain rasio*). Sebelum menghitung rasio perolehan, perlu dilakukan perhitungan nilai informasi dalam satuan bits dari suatu kumpulan objek, yaitu dengan menggunakan konsep entropy untuk membentuk pohon keputusan. Data kemudian dihitung menggunakan algoritma sesuai dengan metodenya kemudian dicari hasil akurasinya. Ada beberapa tahap dalam membentuk pohon keputusan dengan algoritma C.45 antara lain :

1. Menyiapkan data training, dimana data tersebut akan diklasifikasikan.
2. Menentukan akar dari pohon, akar akan diperoleh dari atribut yang terpilih dengan cara menghitung nilai gain dari masing-masing atribut. Nilai gain tertinggi akan dijadikan akar pertama dalam pohon keputusan. Sebelum menghitung nilai gain, hitung dulu nilai entropy dengan persamaan sebagai berikut:

$$Entropy\ (S) = \sum_{i=1}^{n} -pi \times \log_2 pi$$

3. Kemudian hitung nilai *gain* dengan persamaan sebagai berikut :

$$Gain\ (S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} Entropy(Si)$$

4. Untuk langkah 2 hinggga semua record terpartisi.
5. Proses partisi akan berhenti saat :
   a. Semua record pada simpul N mendapat kelas yang sama.
   b. Tidak ada atribut didalam record yang akan dipartisi lagi.
   c. Tidak ada record di dalam cabang yang kosong.

Algoritma K-NN

Metode -Nearest Neighbor (KNN) K-Nearest Neighbor (KNN) adalah metode melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Metode ini bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan training sample. Diberikan suatu titik query, selanjutnya akan ditemukan sejumlah K objek atau titik training yang paling dekat dengan titik query. Nilai prediksi dari query akan ditentukan

K Nearest Neighbor (KNN) adalah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Untuk pemilihan atribut terdiri dari n neighbors (biasa disebut k). parameter k pada testing ditentukan berdasarkan nilai k optimum pada saat training. Nilai k optimum diperoleh dengan mencoba-coba. Menghitung kuadrat jarak euclid (euclidean distance) masing-masing obyek terhadap data sampel yang diberikan.

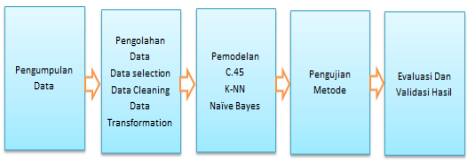
$$D(a,b) = \sqrt{\sum_{k=1}^{d} (a_k - b_k)^2}$$

Berdasarkan rumus 4 dimana matriks D (a,b) adalah jarak skalar dari kedua vektor a data latih) dan b (data uji) dari matriks dengan ukurand dimensi. Mengurut kanhasilobjek-objek tersebut ke dalam kelompok yang mempunyai jarak euclidian terkecil. Mengumpulkan kategori y (klasifikasi nearest neighbor berdasarkan nilai k) Dengan menggunakan kategori nearest neighbor yang paling mayoritas, maka dapat diprekdisikan kategori objek tersebut.

II. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Dalam melakukan analisis dan mencari pola data *Computer Network Intrusion detection System* agar memudahkan penelitian dan dapat berjalan dengan sistematis dan memenuhi tujuan yang diinginkan maka dibuat langkah – langkah dalam tahapan penelitian yang akan dilakukan berikut:



Gambar 1 Tahap Penelitian

2.2 Pengolahan Data Awal

Penelitian ini menggunakan algoritma C.45, Naïve Bayes, K-NN, maka perlu dilakukan pengolahan data untuk mendapatkan dataset yang sesuai dan diinginkan. Data yang akan dijadikan dataset dalam penelitian ini adalah *Computer Network Intrusion detection System*, data yang didapatkan dalam bentuk file spreadsheet berformat excel sehingga tidak dapat langsung digunakan dikarenakan data masih terpisah dan tidak teratur pada beberapa kolom dan sheets. Data yang diperoleh merupakan data *Computer Network Intrusion detection System* berdasarkan data yang didapat dari KDD-CUP99 yang akan dijadikan dataset dengan melalui proses tahapan pengolahan data dengan langkah berikut.

III. ANALISA DAN PEMBAHASAN

3.1 Hasil Pengujian Data

Pengujian data dalam *klasifikasi* serangan pada *Intrusion Detection System (IDS)*. menggunakan metode C4.5, Naive Bayes dan K-NN yang bertujuan untuk mengetahui apakah dalam *klasifikasi* serangan dapat mengtahui hasil nilai *accuracy, recall dan precision* sehingga dapat mengetahui pengetahuan dan informasi dalam meminimalisasi

resiko terhadap pengguna. Pengujian ini dilakukan dengan cara melakukan pengolahan data, 5000 dataset, 4500 data *training* dan data *testing* 500, data dapat di *klasifikasi* sehingga hal ini bertujuan untuk melihat apakah metode yang digunakan dapat *mengklasifikasikan* serangan pada *Intrusion Detection System (IDS).*

Model yang didapatkan dari tiga algoritma *K-NN, C4.5* dan *Naïve Byes* kemudian dilakukan pengujian meng-gunakan *K-fold cross validation*, data yang digunakan dibagi secara acak ke dalam k subset yaitu *D1*,*D2*, *D3* ,*D4*, *D5*, *D6*, *D7*, *D8* … ,*D*10 dengan ukuran yang sama. Dataset akan dibagi menjadi data *training* dan data *testing*. Proses *training* dan *testin*g dilakukan sebanyak 10 kali secara berulang-ulang. Pada iterasi ke-1, partisi *D*1 disajikan sebagai data testing dan partisi sisanya digunakan secara bersamaan dan berurutan sebagai data *training*. Iterasi kedua, subset *D*1,2,…,*Dk* akan dites pada *D*2, Iterasi ketiga, subset *D*1,*D*2, *D3* … ,*Dk* akan dites pada *D3*, dan selanjutnya hingga *D*10). Tabel 4.1 berikut adalah contoh ilustrasi 10 kali pengujian dan tabel 4.2 data *training* dan *testing*.

Tabel 1.Pengujian

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1	500	500	500	500	500	500	500	500	500	500
C2	500	500	500	500	500	500	500	500	500	500
C3	500	500	500	500	500	500	500	500	500	500
C4	500	500	500	500	500	500	500	500	500	500
C5	500	500	500	500	500	500	500	500	500	500
C6	500	500	500	500	500	500	500	500	500	500
C7	500	500	500	500	500	500	500	500	500	500
C8	500	500	500	500	500	500	500	500	500	500
C9	500	500	500	500	500	500	500	500	500	500
C10	500	500	500	500	500	500	500	500	500	500

Tabel 1 Data Training Dan Testing

10 Testing	*Data training*	*Data Testing*
C1	4500	500
C2	4500	500
C3	4500	500
C4	4500	500
C5	4500	500
C6	4500	500
C7	4500	500
C8	4500	500
C9	4500	500
C10	4500	500

Berdasarkan tabel 1 ditunjukkan bahwa nilai *fold* yang digunakan adalah *10-fold cross validation*. Berikut diberikan langkah-langkah pengujian data dengan *10-fold cross validation*.

1. Dataset 5000 yang digunakan dibagi menjadi 10 bagian, yaitu *D*1, *D*2,3, dan *D*4. D5. D6. D7. D8. D9. D10 , *t* = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) digunakan sebagai data testing 500 dan dataset lainnya sebagai data training 4500.
2. Tingkat *accuracy* dihitung pada setiap iterasi (iterasi-1, iterasi-2, iterasi-3, iterasi-4, iterasi-5, iterasi-6, iterasi-7, iterasi-8, iterasi-9, iterasi-10) kemudian dihitung rata-rata tingkat *accuracy* dari seluruh iterasi untuk mendapatkan tingkat *accuracy* data keseluruhan
3. Data tersebut disimpan dalam format *excel workbook* yang selanjutnya diubah menjadi data frame dengan perintah *read excel*.

Berikut Ini adalah data testing untuk di olah ke dalam tools *Rapid Miner*.

3.2 Proses Pengujian Data *(Rapid Miner)*

Dari dataset 5000, data training 4500 dan testing yang diutarakan yaitu sebanyak 500 data, kemudian hasil dari data tersebut menyatakan tingkat *accuracy, Recall* dan *Precision* dalam

*klasifikas*i data *Computer Network Intrusion detection System* pada *Intrusion Detection System (IDS)* dengan algoritma *K-NN, C4.5* dan *Naïve Bayes* Berikut tabel dari keselurahan data yang telah diuji dengan *rapid miner*:

Tabel 3 Hasil Nilai *Accuracy*

algorithm	K-NN	C4.5	Naïve Bayes
10 Testing	Accuracy	Accuracy	Accuracy
C1	83.60%	96.80%	99.00%
C2	80.60%	96.20%	95.00%
C3	85.80%	97.00%	97.00%
C4	83.20%	98.60%	97.40%
C5	81.20%	96.40%	96.20%
C6	81.60%	98.40%	97.70%
C7	84.20%	99.80%	97.20%
C8	82.80%	98.40%	97.00%
C9	82.40%	98.60%	95.80%
C10	80.40%	97.80%	96.80%

Tabel 4 Hasil Nilai Precision

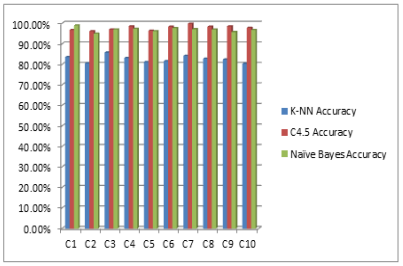
algorithm	K-NN	C4.5	Naïve Bayes
10 Testing	Precision	Precision	Precision
C1	86.00%	94.57%	98.13%
C2	80.39%	95.33%	92.83%
C3	90.79%	99.60%	99.22%
C4	86.69%	97.95%	97.25%
C5	84.05%	95.31%	95.29%
C6	85.54%	97.76%	97.34%
C7	86.77%	99.63%	96.69%
C8	86.11%	97.83%	97.07%
C9	79.15%	100.00%	95.67%
C10	75.58%	98.06%	95.31%

Tabel 5 Hasil Nilai Recall

algorithm	K-NN	C4.5	Naïve Bayes
10 Testing	Recall	Recall	Recall
C1	82.06%	99.62%	100.00%
C2	81.35%	97.22%	97.62%
C3	81.58%	94.74%	95.11%
C4	83.68%	99.65%	98.26%
C5	80.30%	98.14%	97.77%
C6	78.41%	99.24%	96.97%
C7	83.21%	100.00%	98.13%
C8	79.78%	99.26%	97.43%
C9	88.54%	97.23%	96.05%
C10	78.47%	96.65%	97.13%

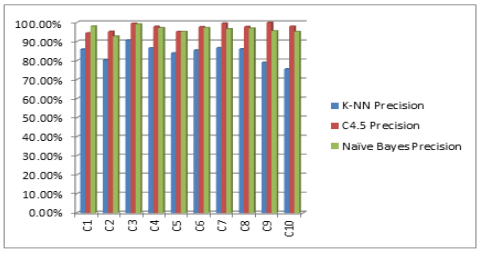
Dari data yang telah diuji, kemudian hasil dari data tersebut menyatakan tingkat *accuracy* , *Recall* dan *Precision* dalam *klasifikasi* data dengan algoritma *K-NN, C4.5* dan *Naïve Bayes*. Berikut grafik dari keselurahan data yang telah di uji dengan rapid mener:

1. Accuracy
   Grafik hasil dari nilai accuracy dengan hasil uji data training dan testing.



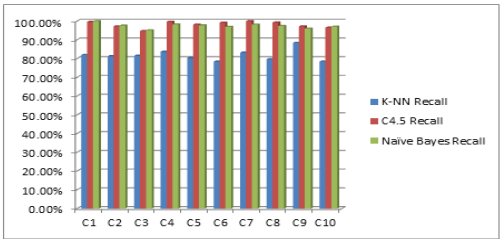
Gambar 2 Grafik Accuracy

2. Precision
   Grafik hasil dari nilai precision dengan hasil uji data training dan testing.



Gambar 3 Grafik Precision

3. Recall
   Grafik hasil dari nilai recall dengan hasil uji data training dan testing.



Gambar 4 Grafik Recall

Dari dataset sebanyak 5000, data *training* 4500 dan hasil data *testing* 500 yang kemudian dilakukan 10 kali pengujian setiap 500 atau setiap partisi data testing dan dari data tersebut diuji dangan algoritma *K-NN, C.45 dan Naïve Bayes*.

Menyatakan *accuracy* 82.58%, *Recall* 81.73% dan *Precision* 84.11 %

dalam *klasifikasi* data *Computer Network Intrusion detection System* dari algoritma *K-NN*

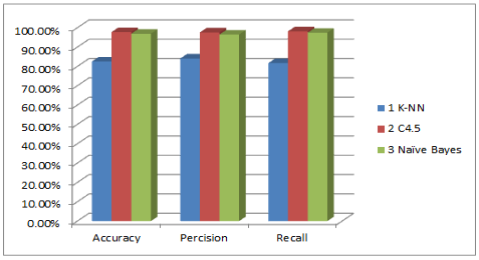
Menyatakan *accuracy* 97.80%, *Recall* 98.18% dan *Precision* 97.60 % dalam *klasifikasi* data *Computer Network Intrusion detection System* dari algoritma C.45*:*

Menyatakan *accuracy* 96.91% *,Recall* 97.45% *dan Percision 96.48 %* dalam *klasifikasi* data *Computer Network Intrusion detection System* dari algoritma *Naïve bayes.*

Tabel 6 Nilai Rata – Rata Accuracy, Precision Dan Recall

No	Algoritma	Accuracy	Percision	Recall
1	K-NN	82.58%	84.11%	81.74%
2	C4.5	97.80%	97.60%	98.18%
3	Naïve Bayes	96.91%	96.48%	97.45%

Dari hasil pengujian dengan melakukan 10 kali pengujian secara random dengan mengahasilkan tingkat *accuracy, precision* dan *recall* teringgi adalah *C.45*



Gambar 5 Grafik Nilai Rata – Rata Accuracy, Precision Dan Recall

3.3 Analisis Hasil Pengujian

Hasil dari penelitian ini merupakan pengujian data *Intrusion Detection System (IDS*) menggunakan algoritma *K-NN, Naïve Bayes* dan *C.45* menggunakan dataset 5000 data dibagi dua testing data 10% dan training data 90% dalam pengujianya secara random dari data C = (1,2,3,4,5,6 7……...10) 10 kali pengujian menghasilkan hasil yang berbeda-beda.

Dari pengujian beberapa data dan atribut, K-NN menghasilkan recall 81.73% dan accuracy 82.58% , Precision 84.11 % yang lebih rendah sedangkan algoritma C.45 memiliki nilai accuracy 97.80% , Recall 98.18 dan Precision 97.60 % Hal ini disebabkan algoritma K-NN dalam suatu label, kelas dapat mencari hasil yang terdekat. Dari hasil pengujian data dalam memiliki 8 atribut yaitu atribut (attack), lebel normal, dos, probe, r2l. Hal ini disebabkan atau normal dianggap yes (adanya serangan) yang seharus No (Tidak ada serangan) jadi memiliki nilai *accuracy, precision* dan recall terendah disebabkan K-NN dapat mencari dan menidentifikasi kelas atau atribut terdekat dari sebuah data, training data, testing data memiliki kedekatan jarak, data testing sebagai dasar pengklasifikasian.

Sedangkan algoritma C.45 menghasilkan hasil optimal, berdasarkan hasil yang sudah didapatkan dalam penelitian ini maka algoritma yang menghasilkan tingkat accuracy 97.80%, Recall 98.18 dan Precision 97.60 % yang paling tinggi yaitu C.45 karena atribut (attack) normal, dos, probe dan r2l, normal (tdak ada serangan), yes (adanya serangan) setiap data berdasarkan hasil dari information gain tinggi dalam penentuan label atau atribut mendapatkan hasil yang baik dalam suatu data atau semua atribut dalam meningkatkan hasil accuracy, recall dan precision. Meningkatnya akurasi ini dapat mempermudah dalam pengambilan keputusan dan upaya pencegahan dari setiap perbedan algoritma merupakan salah satu faktor yang menyebabkan nilai akurasi tinggi karena setiap atribut dan kelas atau label

memiliki pengaruh pada algoritma tersebut pada penelitian ini terdapat pengujian secara random (cross validation), C45, KNN dan Naive Bayes. Selain itu ada beberapa faktor yang mempengaruhi nilai accuracy, precision dan recall ketepatan relevansi dalam pemilihan algoritma dan semakin relevan atribut, kelas atau label yang terseleksi maka nilai akurasi yang didapatkan semakin tinggi.

IV. KESIMPULAN

4.1 Kesimpulan

Hasil dari pengujian algoritma K-NN, Naïve Bayes, C.45 bahwa hasil dari pengujian dapat di simpulkan sebagai berikut: Hasil pegujian dari algoritma Naïve Bayes, K-Nearest Neighbor dan C.45, C.45 mendapatkan hasil lebih baik dari tingkat recall dan precision, accuracy. Dari hasil pengujian data dalam memiliki 8 atribut. Pada suatu data atau atribut (attack) terdapat lebel normal, dos, probe, r2l. dapat menyimpulkan hasil yang rendah karena disebabkan atau normal di anggap yes (adanya serangan) yang seharus No (Tidak ada serangan), sedangkan algoritma C,45, atribut (attack) normal, dos, probe dan r2l, normal (tdak ada serangan), yes (adanya serangan) sehingga menghasilakn tingkat accuracy 97.80%, recall 98.18% dan 97.60% paling optimal dalam pengujian data

DAFTAR PUSTAKA

Dicky Nofriansyah, Gunadi Widi Nurcahyo, Penerapan Data Mining, Jogyakarta 2015

Jupriyadi, (2018). Implementasi Seleksi Fitur Menggunakan Algoritma FVBRM Untuk Klasifikasi Serangan Pada Intrusion Detection System (IDS)

Khaerani & Handoko, (2015). Klasifikasi Serangan pada Intrusion Detection System (IDS) Dengan Algoritma C.45.

Kharisma Muchammad (2016). Deteksi Intrusi dengan Jumlah Jarak dari Centroid dan Sub-centroid.

Kusrini, Emah Taufiq Luthfi (2009). Algortima Data Mining. Andi Yogyakarta

Muhammad Satria Nugraha, (2010). Implementasi Intrusion Detection System (IDS) Untuk filtering Paket Data. Implementasi dan Analisa Hasil Data Mining

Oktavia Ari Marlita, Adiwijaya, Angelina Prima Kurniati, (2015). Anomaly Detection pada Intrusion Detection System (IDS) Menggunakan Metode Bayesian Network.

Osiris Villacampa, (2015).Feature Selection and Classification Methods for Decision Making: A Comparative Analysis.

Retno Tri Vulandari, 2017 . Data Mining. Gava Media.

Silalahi, Kristiani Desri., Murfi, Hendri., Satria, Yudi. (2017). Studi Data Mining. Informatika

....., Perbandingan Pemilihan Fitur untuk Support Vector Machine pada Klasifikasi Penilaian Risiko Kredit, 1(2), 119–136.