IMPLEMENTASI METODE NAIVE BAYES UNTUK INTRUSION DETECTION SYSTEM (IDS)

Arief Prasetyo[1], Luqman Affandi [2], Dedi Arpandi [3]

[1,2]Jurusan Teknologi Informasi, Program Studi Teknik Informatika,[3] Politeknik Negeri Malang
tiyok_pras@yahoo.com [1], laffandi@polinema.ac.id [2] , dediarpandi@gmail.com[3]

Abstrak

IDS berfungsi untuk mengidentifikasi *traffic* atau lalu-lintas data pada sebuah jaringan komputer dimana IDS dapat menentukan apakah *traffic* aman, mencurigakan atau bahkan terindikasi merupakan serangan. Permasalahan muncul ketika ada aktifitas-aktifitas yang mencurigakan atau bahkan aktifitas tersebut merupakan serangan namun tidak terdaftar pada *rule* atau aturan yang diinputkan sehingga hal itu sangat membahayakan sebuah jaringan komputer. Tujuan dari penellitian ini adalah membangun sistem deteksi pola serangan baru menggunakan metode *naive bayes* untuk mengatasi serangan-serangan baru yang muncul, dan yang belum terdaftar pada *signature* serta untuk meningkatkan akurasi pendeteksian serangan-serangan baru pada *Intruison Detection System* (IDS). Data yang digunakan pada penelitian ini adalah data NSL-KDD, NSL-KDD telah menyediakan *data training* dan *data testing* untuk proses penelitian klasifikasi serangan. Dari data NSL-KDD akan dilakukan klasifikasi serangan menggunakan metode *naive bayes* agar serangan-serangan baru dapat terklasifikasi. Penelitian yang menggunakan metode *naive bayes* ini telah berhasil melakukan klasifikasi serangan-serangan baru dengan akurasi kebenaran adalah sebesar 81-84,67 %.

Kata kunci : *Intruison Detection System* (IDS), NSL-KDD, *Naive Bayes*

1. Pendahuluan

Kemanan adalah salah satu aspek yang penting dalam masalah internet khususnya jaringan komputer. Sebuah jaringan komputer harus mampu memberikan rasa aman terhadap akses yang dilakukan oleh seorang user, dengan memberikan jaminan informasi atau data pribadi aman dari pengaksesan seorang intruder (penyerang).

Namun dari tahun ke tahun serangan terhadap jaringan komputer khususnya internet mengalami peningkatan. Berdasarkan laporan dari Kaspersky Lab jumlah serangan melalui browser internet sejumlah 23.680.646 pada tahun 2007, meningkat menjadi 73.619.767 pada tahun 2009 dan meningkat lagi menjadi 580.371.937 pada tahun 2010. Internet browser menjadi alat utama dalam menyebarkan program-program malicious diantara sebagian besar pengguna komputer pada tahun 2010. Algoritma Kaspersky Security Network (KSN) hanya mampu mendeteksi serangan web sebesar 60 % (Gostev, A, 2010).

Salah satu upaya melindungi jaringan dari ancaman-ancaman intruder (penyerang) adalah membangun sistem deteksi intrusi atau *Intrusion Detection System* (IDS). *Intrusion detection* adalah proses memonitor kejadian pada sistem komputer atau jaringan dan menganalisanya untuk memberikan tanda insiden yang mungkin, yang mana yang merupakan pelanggaran atau mendekati pelanggaran sebuah kebijakan keamanan komputer, kebijakan penggunaan yang disetujui atau praktik keamanan standar (Scarfone, K, 2007). IDS berfungsi untuk mengidentifikasi *traffic* atau lalu-lintas data pada sebuah jaringan komputer dimana IDS dapat menentukan apakah *traffic* aman, mencurigakan atau bahkan terindikasi merupakan serangan.

Permasalahan muncul ketika ada aktifitas-aktifitas yang mencurigakan atau bahkan aktifitas tersebut merupakan serangan namun tidak terdaftar pada rule atau aturan yang diinputkan sehingga hal itu sangat membahayakan sebuah jaringan komputer. Oleh karena itu dibutuhkan sebuah sistem klasifikasi serangan yang berfungsi untuk mengklasifikasi traffic jaringan yang ada dan dari klasifikasi tersebut akan diketahui apakah sebuah aktifitas pada sebuah traffic jaringan tersebut serangan atau bukan serangan. Dari hasil klasifikasi tersebut juga dapat digunakan menjadi dasar untuk membuat rule baru yang akan diinputkan menjadi aturan-aturan pada aplikasi IDS yang digunakan.

1.1 Rumusan Masalah

Bagaimana menerapkan metode naive bayes untuk mengklasifikasi serangan yang mampu

mengidentifikasi serangan-serangan baru yang belum terdaftar pada signature.

1.2 Tujuan

Membangun sistem deteksi pola serangan baru menggunakan metode *naive bayes* untuk mengatasi serangan-serangan baru yang muncul, dan yang belum terdaftar pada *signature*.

1.3 Batasan Masalah

Pada penelitian ini ditentukan beberapa batasan masalah, yaitu sebagai berikut :

1. Data Training dan Data Testing berasal dari NSL-KDD (nsl.cs.unb.ca/NSL-KDD/).
2. Metode yang digunakan pada klasifikasi serangan adalah metode naive bayes.
3. Sistem ini dibuat untuk membantu adminintrator jaringan.

2. Tinjauan Pustaka

2.1. Keamanan Jaringan

Komputer yang terhubung dengan jaringan memiliki resiko ancaman keamanan lebih besar daripada komputer yang tidak terhubung dengan jaringan. Dengan beberapa cara, jaringan komputer dapat lebih dioptimalkan dari resiko ancaman pihak yang tidak memiliki hak akses terhadap sumber yang ada pada jaringan tersebut, namun hal tersebut akan berbanding terbalik dengan kenyamanan akses pengguna, dimana tingkat keamanan yang tinggi akan membuat pengguna tidak nyaman, sedangkan tingkat keamanan yang rendah maka akses semakin nyaman.

2.2. *Intrusion Detection System*

IDS merupakan sistem yang dapat melihat pola dari serangan-serangan yang terdapat pada jaringan komputer, pola tersebut berupa paket yang lewat yang teridentifikasi oleh IDS sebagai paket yang mengandung pola serangan. Terdapat dua katagori IDS yaitu Network Based IDS dimana jenis ini dapat menganalisa semua paket didalam jaringan dan yang kedua disebut client based IDS yang dapat menganalissa logfile yang berisi pola mencurigakan dari sebuah serangan terhadap suatu client yang lewat pada sebuah jaringan (Rafiudin, R, 2012).

2.3. *Naive Bayes*

Menurut (Witten, I.H, 2011) algoritma naive bayes merupakan algoritma yang menggunakan pendekatan statistik dalam mengambil keputusan. Algoritma naive bayes berdasarkan teorema bayes bahwa semua atribut memberikan kontribusi yang sama penting dan saling bebas pada kelas tertentu.

Walaupun teori ini bertentangan dengan kehidupan nyata bahwa atribut tidak sama penting atau independen, tetapi naive bayes menunjukkan performa yang mampu bersaing dengan algoritma klasifikasi yang terkenal, decision tree dan neural network (Gostev, A, 2010).

2.4. My SQL

Mysql adalah salah satu aplikasi manajemen database SQL . aplikasi ini bersifat gratis dan opensource , dalam database mysql terdapat beberapa penggunaan data dan semuanya dapat terorganisir. Mysql tidak hanya tersedia pada sistem operasi unix atau linux, mysql juga dapat digunakan diatas platform windows.

Mysql mempunyai beberapa fasilitas dan kelebihan dibandingkan jenis manejemen database lainnya, untuk itu peneliti menggunakan Mysql sebagai aplikasi penyimpanan data. Diantaranya kelebihan itu adalah :

• MySQL dapat berjalan stabil pada berbagai sistem operasi seperti Windows, Linux,FreeBSD, Mac Os X Server, Solaris, Amiga, dan sistem operasi lainnya.

• MySQL didistribusikan sebagai perangkat lunak sumber terbuka, dibawah lisensi GPL sehingga dapat digunakan secara gratis.

• MySQL dapat digunakan oleh beberapa pengguna dalam waktu yang bersamaan tanpa mengalami masalah.

• *MySQL* memiliki kecepatan yang sangat baik dalam menangani query sederhana, dengan kata lain dapat memproses lebih banyak SQL per satuan waktu.*MySQL* memiliki ragam tipe data yang sangat kaya, seperti *signed* / *unsigned integer*, *float*, *double*, *char*, *text*, *date*, *timestamp*, dan lain-lain.

• Perintah dan Fungsi. *MySQL* memiliki operator dan fungsi secara penuh yang mendukung perintah *Select* dan *Where* dalam perintah (*query*).

• *MySQL* memiliki beberapa lapisan keamanan seperti level *subnetmask*, nama *host*, dan izin akses *user* dengan sistem perizinan yang mendetail serta sandi terenkripsi.

2.5. PHP

PHP adalah sebuah bahasa pemrogaman yang dapat berfungsi untuk membuat sebuah website ataupun aplikasi web yang sifatnya dinamis. Kelebihan dari PHP adalah dapat berinteraksi dengan database seperti mysql, sehingga perintah SQL dapat dijalankan melalui pemrogaman PHP. PHP disebut juga sebagai bahasa scripting sehingga semua proses dijalankan pada sisi server.

Peneliti menggunakan bahasa pemrograman PHP karena pada penelitian ini hasil klasifikasi serangan akan ditampilkan melalui web sehingga bahasa pemrograman PHP adalah bahasa yang paling tepat untuk digunakan pada penelitian ini. Selain itu

PHP juga dapat langsung berinterkasi dengan mysql, sehingga mempermudah dalam menyelesaikan penelitian ini.

2.6. Dataset NSL-KDD

Pada penelitian digunakan dataset NSL-KDD sebagai data penelitian yang memiliki data serangan yang lengkap baik data training dan data testing. NSL-KDD menghasilkan empat kategori serangan yang sering terjadi yaitu :

• 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.

• PROBING (berusaha mencari kelemahan sistem yang ada), misal : insidesniffer, Ipsweep, ls_domain, Mscan, NTinfoscan, Nmap, queso, resetscan, Saint, Satan.

• R2L (Remote To Local - 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.

• U2R (User To Root - 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.

2.7. Format Dataset KDD '99

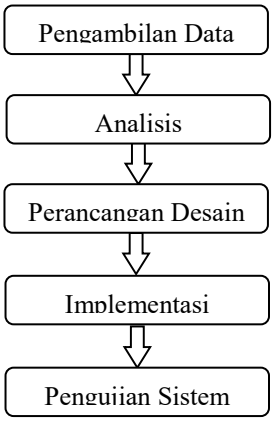
KDDCUP '99 dataset (kdd.ics.uci.edu) merupakan data hasil preprocessing yang berbasis pada data DARPA 1998 yang disediakan untuk perancangan sistem pendeteksian intrusi yang mana digunakan untuk melakukan evaluasi metodologi yang berbeda dari pendeteksian intrusi. Pada tahun 1999 dilakukan preprocessing pada data tcpdump ini untuk dimanfaatkan dalam pendeteksian intrusi pada kegiatan "International Knowledge Discovery and Data Mining Tools Competition". Atribut yang dihasilkan pada KDDCUP '99 terdiri dari 41 atribut ditambah 1 untuk label.

3. Metodologi Penelitian

3.1. Metodologi

Dalam pelaksanaan dan pengimplementasian sistem deteksi serangan menggunakan metode naive bayes, pengembangan aplikasi yang digunakan adalah waterfall.

Gambar 1 merupakan tahapan – tahapan yang dilakukan pada sistem deteksi serangan, terdapat 5 tahapan yang dilaksanakan untuk membangun sistem yaitu sebagai berikut.



Gambar 1. Diagram Pengembangan Aplikasi

4. Analisis dan Perancangan

4.1. Pengambilan Data

Pada tahap ini melakukan pengumpulkan data dan informasi dengan cara membaca referensi dari buku maupun dari internet, khususnya referensi-referensi yang terkait dengan IDS, serta melakukan wawancara kepada narasumber yang terkait untuk menambah wawasan tentang klasifikasi serangan pada IDS. Penelitian ini akan menggunakan data NSL-KDD sebagai *data testing* dan *data training*, data NSL-KDD dapat diperoleh di nsl.cs.unb.ca/NSL-KDD/.

4.2. Analisis Permasalahan

Pada tahap ini dilakukan analisis permasalahan yang berasal dari pengambilan data baik studi literatur maupun wawancara kemudian permasalahan yang ada akan dijadikan landasan dan acuan dalam membangun sistem ini. Beberapa permasalahan yang telah diselesaikan yaitu :

1. Sistem klasifikasi membutuhkan data *training* dan *testing* yang diperoleh dari *raw data KDDCUP '99*
2. Sistem membutuhkan *preprocessor* yang digunakan untuk menerapkan metode dalam melakukan klasifikasi serangan
3. Atribut yang digunakan adalah *src_bytes, dst_bytes, count, srv_count, dst_host_count, dst_host_srv_count, dst_host_same_src_port_rate,* dan *dst_host_srv_diff_host_rate.*
4. Membutuhkan *web interface* yang digunakan untuk penyajian hasil dari klasifikasi serangan.
5. Aplikasi dibuat untuk membantu administrator jaringan dalam mengamankan sistem jaringan terhadap serangan-serangan baru.

4.3. Perancangan Desain dan Sistem

Tahap perancangan desain dan sistem akan digambarakan dengan DFD (*Data Flow Diagram*), Rancangan *Database* dan *Flowchat*, untuk mempermudah membaca dan menerpakan sistem yang digunakan.

4.4. Implementasi

Pada tahap ini dilakukan implementasi sistem yang terdiri dari membuat *data training* dan *data testing*, menerapkan metode *naive bayes*, membuat *web interface*.

Sistem menggunakan data NSL-KDD sebagai data mentah, NSL-KDD telah menyediakan *data training* dan *data testing* untuk penelitian terkait dengan *Intrusion Detection System* (IDS). *Raw data* yang diperoleh dari NSL-KDD kemudian diubah menjadi format KDDCUP '99 yaitu sebanyak 41 field atau fitur, dari 41 *field* tersebut dipilih 8 *field* yang akan digunakan sebagai acuan data training dan data testing. 8 *field* tersebut yaitu, *src_bytes, dst_bytes, count, srv_count, dst_host_count, dst_host_srv_count, dst_host_same_src_port_rate,* dan *dst_host_srv_diff_host_rate*.

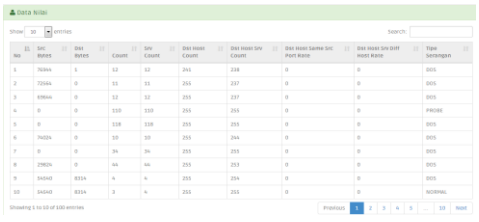
4.5. Pengujian Sistem

Menguji sistem klasifikasi serangan menggunakan metode *naive bayes* yang bertujuan untuk mengetahui apakah sistem yang dibuat dapat telah dapat mengklasifikasi serangan dan dapat membedakan jenis-jenis serangan.

Pengujian sistem dilakukan dengan cara melakukan *testing* terhadap metode dengan menginputkan data baru yang tidak terdapat dalam data klasifikasi sehingga hal ini bertujuan untuk melihat apakah metode yang digunakan dapat mengklasifikasikan serangan-serangan baru dengan acuan pola-pola serangan yang sudah ada sebelumnya.

5. Implementasi Sistem

5.1. Tampilan hasil klasifikasi menggunakan metode *naive bayes*



Gambar 2. Penerapan Metode *Naive Bayes*

6. Pengujian dan Pembahasan

6.1 Pembahasan

Salah satu pengukur kinerja klasifikasi adalah tingkat akurasi. Sebuah sistem dalam melakukan klasifikasi diharapkan dapat mengklasifikasi semua set data dengan benar, tetapi tidak dipungkiri bahwa kinerja suatu sistem tidak bisa 100% akurat. Untuk pada penenlitian ini akan dilakukan uji akurasi yang berfungsi melihat seberapa baik akurasi klasifikasi serangan menggunakan metode naive bayes. Berikut adalah formula untuk menghitung akurasi.

Akurasi= (jumlah data benar)/(jumlah data) x 100 %

Data training yang akan digunakan pada pengujian akurasi ini adalah sebanyak 1500, 3000, 5000 data sedangkan untuk *data testing* sebanyak 100, 150, 200 data. Masing-masing *data testing* akan diuji pada masing-masing *data training*, berikut tabel 1 adalah tabel rincian pembagian data yang digunakan pada penelitian ini.

Tabel 1 Rincian Data Pengujian Akurasi

No	Jumlah Data Training	Jumlah Data Testing
1	1500 data	100 data
		150 data
		200 data
2	3000 data	100 data
		150 data
		200 data
3	5000 data	100 data
		150 data
		200 data

Setelah dilakukan uji coba akurasi menggunakan data training sebanyak 1500, 3000, 5000 data serta menggunakan data testing sebanyak 100, 150, 200 data, hasil akurasi menunjukkan bahwa semakin banyak data training maka hasil akurasi kebenaran akan semakin baik. Berikut tabel 2 perbandingan banyaknya data training dan hasil akurasi yang diperoleh.

Tabel 2 Perbandingan Hasil Akurasi *Data Testing*

No	Jumlah Data Testing	Jumlah Data Training	Akurasi
1	100 data	1500 data	81 %
		3000 data	82 %
		5000 data	84 %
2	150 data	1500 data	81,33 %
		3000 data	82,67 %
		5000 data	84,67 %
3	200 data	1500 data	82,5 %
		3000 data	83 %
		5000 data	84 %

Dari tabel 1 dapat dilihat bahwa semakin banyak data training yang digunakan dalam proses klasifikasi serangan maka akurasi kebenaran dalam menentukan serangan akan semakin baik. Pada penelitian klasifikasi serangan menggunakan metode naive bayes ini telah sesuai dengan apa yang diharapkan yaitu dapat mengklasifkasikan serangan-serangan baru dengan acuan data training yang ada.

7. Kesimpulan

7.1. Kesimpulan

Dari beberapa uji coba yang didapatkan dari penelitian ini, dapat disimpulkan beberapa hal sebagai berikut :

1. Metode naive bayes dapat digunakan sebagai klasifikasi serangan-serangan baru pada Intruison Detection System (IDS).
2. Semakin banyak data training yang digunakan maka hasil dari akurasi kebeneran akan semakin baik.
3. Akurasi kebanaran pada klasifikasi serangan menggunakan metode naive bayes adalah 81-84,67 %.

7.2. Saran

Dari apa yang telah dilakukan dan diujicobakan dalam penelitian ini, dari segi akurasi yang telah didapatkan sudah menunjukkan hasil positif. Namun untuk lebih menyempurnakan penelitian ini, mungkin diperlukan beberapa masukan dari peneliti yaitu :

1. Menggunakan *field-field* tambahan untuk proses klasifikasi serangan memungkinkan untuk dapat meningkatkan hasil keakuratan.
2. Menggunakan metode-metode klasifikasi lain dan membandingkan tingkat akurasi kebenaran dalam melakukan klasifikasi serangan.

Daftar Pustaka:

Darujati, C. (2010). "*Perbandingan Klasifikasi Dokumen Teks Menggunakan Metode Naive Bayes dab K-Nearest Neighbor*" .

Gostev, A. & Namestnikov, Y., 2011. *"Kaspersky Security Bulletin 2010".* Statistics, 2010. [Online] Available at:http://www.securelist.com/en/analysis/204792162/Kaspersky_Security_Bulletin_2010_Statistics_2010 [Accessed 28 Desember 2016].

Mitchell, T., 1997. "*Machine Learning".* New York: McGraw Hill.

Novi Anisyah, M. Z. (2011). "*APLIKASI MOBILE UNTUK K-NEAREST NEIGHBOR PADA INTRUSION DETECTION SYSTEM BERBASIS SNORT*".

Prasetyo, E. 2012. *"Data Mining Konsep dan Aplikasi Menggunakan Matlab".* Yogyakarta : Andi.

Rafiudin, R. 2012. *"Mengganyang Hacker Dengan Snort*". Yogyakarta:Penerbit Andi

Santoso, D. (2013). *"Perbandingan Kinerja Metode Naive Bayes, K-Nearest Neighbor dan Metode Gabungan K-Means dan LVQ dalam Pengkategorian Buku Komputer Bahasa Indonesia Berdasarkan Judul dan Sinopsis".*

Scarfone, K. & Mell, P., Februari, 2007. Special Publication 800-94: *"Guide To Intrusion Detection and Prevention Systems. Gaithersburg, Maryland: National Institute Standard and Technology".*

Sofana, I. 2012. *"Cisco CCNA dan Jaringan Komputer"*. Bandung: Penerbit Informatika.

Tiyas, F. I., Hadi, M. Z., & K, E. M. (2011). "*APLIKASI WEB UNTUK METODE FUZZY NEUARAL NETWORK PADA INTRUSION DETECTION SYSTEM BERBASIS SNORT".*

Wicaksana, P. D. (2015). *PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR DAN NAIVE BAYES UNTUK STUDI DATA "WISCONSIN DIAGNOSIS BREAST CANCER"*. Yogyakarta.

Witten, I.H., Frank, E. & Hall, M.A., 2011. *"Data Mining Practical Machine Learning Tools and Technique Third Edition*". New York: Morgan Kaufmann.

Ying Yang, G. I. (2002). "*A Comparative Study of Discretization Methods for Naive-Bayes Classifier"s.*