DATA MINING UNTUK KLASIFIKASI SERANGAN JARINGAN MENGGUNAKAN METODE NAIVE BAYES

SKRIPSI

Diajukan Untuk Memenuhi Salah Satu Syarat Penyusunan Skripsi
Pada Fakultas Teknik
Universitas Islam Riau Pekanbaru

DISUSUN OLEH:

SADELA PRANATA
133510527

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS ISLAM RIAU
PEKANBARU
2020

KATA PENGANTAR

Dengan mengucapkan puji syukur kepada Allah SWT yang telah melimpahkan segala rahmat dan hidayah Nya kepada penulis, sehingga penulis dapat menyelesaikan laporan skripsi yang berjudul "**Data Mining Untuk Klasifikasi Serangan Jaringan Menggunakan Metode Naive Bayes** ", yang menjadi salah satu syarat untuk menyelesaikan program studi Teknik Informatika jenjang Strata satu (S1) Universitas Islam Riau (UIR) Pekanbaru. Shalawat serta salam semoga tetap tercurah kepada junjungan Alam Yakni Nabi Besar Muhammad SAW.

Dalam proses pembuatan laporan skripsi ini, penulis menyadari bahwa penulisan laporan skripsi ini masih jauh dari kata sempurna dan banyak mengalami kendala. Namun, dalam penyelesaian laporan ini penulis mendapat banyak sekali bantuan, dorongan dan bimbingan yang sangat berharga yang diberikan kepada penulis, untuk itu pada kesempatan ini penulis ingin mengucapkan rasa terimakasih kepada :

1. Dr. Arbi Haza Nasution, B.IT., M.IT selaku ketua program studi Teknik Informatika.
2. Ibu Ana Yulianti, ST.,M.Kom selaku sekretaris program studi Teknik Informatika
3. Ibu Ir.Des Suryani, M.Sc selaku pembimbing, yang telah banyak membantu saya memberikan pengarahan dan bimbingan dalam menyelesaikan skripsi ini dengan baik.

4. Tata Usaha Fakultas Teknik yang selalu mambantu dalam pembuatan surat dan pengesahan.
5. Semua pihak Provider Internet Wanxp, yang telah membantu menyelesaikan laporan skripsi ini yang tidak bisa penulis sebutkan satu persatu.

Penyusunan laporan skripsi ini telah diusahakan dengan semaksimal mungkin, namun penulis menyadari masih ada kekurangan, penulis mengharapkan kritik dan saran yang membangun dari pembaca agar dapat disempurnakan lagi kemudian hari.

Akhir kata, penulis berharap penyusunan laporan skripsi ini dapat bermanfaat dan dapat dikembangkan lebih lanjut.

Pekanbaru, 03 September 2020

Penulis

ABSTRAK

Penilitian ini bertujuan untuk merancang dan membangun aplikasi klasifikasi serangan jaringan serta mengetahui tingkat akurasi dari metode Algoritma *Naïve Bayes* pada klasifikasi jaringan. Manfaat dari penelitian ini ialah membantu mengklasifikasikan jaringan agar dapat meminimalisir serangan jaringan dengan penanganan berdasarkan klasifikasi yang dihasilkan sehingga lebih mudah dalam penanganan. Memberi pengetahuan kepada penulis mengenai penggunaan metode Naive Bayes untuk klasifikasi jaringan. Metode penelitian deskriptif kualitatif dengan teknik pengumpulan data yaitu melakukan wawancara dengan pada salah satu karyawan provider internet wanxp, lokasi penelitian di jl. Ir. H. Juanda. Hasil penelitian ini mencapai nilai 82% tingkat kesamaan data antara data dariNSL KDD dan sistem sehingga memiliki peforma baik dengan pembagian data training sebanyak 70% yaitu 46873 dan data testing 30% yaitu 18661.

Kata Kunci : Serangan Jaringan, mengklasifikasikan, *Naive Bayes*

ABSTRACT

This study aims to design and build network attack classification applications and determine the level of accuracy of the Algorithm method Naïve Bayes in network classification. The benefit of this research is to help classify networks in order to minimize network attacks by handling based on the resulting classification so that it is easier to handle. Provide knowledge to the author about the use of the Naive Bayes method for network classification. Qualitative descriptive research method with data collection techniques, namely conducting interviews with one of the employees of the internet provider wanxp, the research location at jl. Ir. H. Juanda. The results of this study reached a value of 82%, the level of data similarity between the data from NSL KDD and the system so that it has a good performance by sharing training data as much as 70%, namely 46873 and testing data 30%, namely 18661.

Keywords: *Network attack, classify, Naive Bayes*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I

PENDAHULUAN

I.1 Latar Belakang

Saat ini internet meningkat secara signifikan, meningkatnya penggunaan internet meningkat pula gangguan atau serangan terjadi berupa serangan dari *cracker* yang menemukan kelemahan protokol internet, sistem operasi serta aplikasi. Pada tahun 2015 *Kaspersky Lab* menyatakan bahwa hasil survei untuk ancaman-ancaman serangan dari beberapa sumber dengan persentase terbesar yaitu *browser* 62% di ikuti oleh *android* 14% (Garnaeva, 2015). Ancaman serangan bervariasi tergantung bagaimana jenis keamanan jaringan yang diterapkan. Jaringan dianggap aman jika keamanan jaringan tidak berfokus pada keamanan komputer saja, melainkan saat transmisi data (Daya, 2015).

Adapun upaya untuk pencegahan potensi serangan yang telah dikembangkan yaitu *Intrusion Detection System* (*IDS*). IDS diindikasikan dengan instrumen yang berkembang untuk mendeteksi pelanggaran terhadap keamanan sistem. IDS berfokus pada sistem pertahanan yang mendeteksi kemungkinan masalah dan informasi *logging*. IDS dibagi menjadi dua metode yaitu *misuse-based* dan *anomaly-based.* Model mengeksekusi teknik pencocokan pola sederhana dengan dasar pola serangan pada *database* dan menghasilkan banyak *false* positif rendah disebut *Misuse Detection* sedangkan model *Anomaly Detection* mendeteksi dengan cara mencari abnormal pada jaringan (Jabez, 2015).

Sehingga pendeteksian serangan jaringan dibutuhkan klasifikasi serangan jaringan dikarenakan banyaknya jenis serangan pada jaringan. Perlu dibangun sebuah sistem yang mampu mengklasifikasikan jenis serangan tersebut. Penelitian terkait tentang klasifikasi jaringan diantaranya yaitu penelitian Edi tahun 2017, dilakukan pengklasifikasian menggunakan Algoritma *naïve bayes* dengan dataset NSL KDD untuk mengklasifikasi jenis serangan berupa normal atau anomaly pada IDS. Pada penelitian ini didapatkan hasil akurasi rata-rata 92% dengan seleksi atribut menggunakan *wrapper subsetevalution* (WRP). Proses klasifikasi dapat dilakukan menggunakan metode Algoritma *naïve bayes*. Algoritma *Naïve Bayes* dapat menghasilkan tingkat akurasi hingga 81,89% saja untuk kasus pendeteksian gangguan jaringan komputer dengan menggunakan seleksi atribut berbasis korelasi. (Bekti, 2013).

Pada penelitian ini penulis melakukan perancangan dan membangun sebuah sistem untuk mengklasifikasi serangan pada jaringan komputer dengan dataset NSL KDD tahun 2015 untuk mengatasi rangkaian serangan jaringan. Teknik yang digunakan adalah algoritma *Naïve Bayes*. Berdasarkan permasalah tersebut diatas penulis akan mengangkat judul skripsi dengan judul **"Data Mining Untuk Klasifiksai Serangan Jaringan Menggunakan Metode Naive Bayes".**

I.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, maka dapat di identifikasi masalah dalam pembuatan tugas akhir ini. antara lain:

1. Banyaknya ancaman dari berbagai jalur jaringan sehingga perlu adanya klasifikasi terhadap serangan jaringan.
2. Banyaknya jenis serangan pada jaringan server yang perlu dilakukan identifikasi sehingga serangan pada jaringan dapat diklasifikasi.
3. Belum adanya dilakukan klasifikasi terhadap serangan jaringan dengan banyaknya jenis serangan yang terjadi pada jaringan.

I.3 Perumusan Masalah

Berdasarkan permasalahan yang dikemukakan pada latar belakang, maka rumusan masalah pada penelitian ini adalah bagaimana implementasi Algoritma *Naive Bayes* dalam melakukan klasifikasi terhadap serangan jaringan.

I.4 Batasan Masalah

Adapun batasan masalah dalam penelitian ini agar tidak meluas maka diperlukan batasan sebagai berikut :

1. Data yang digunakan adalah data NSL KDD (*knowledge discovery in database*) yang di peroleh dari provider internet *wanxp*.
2. Data yang digunakan adalah data tahun 2015
3. Tipe pada class penelitian ini yaitu berupa Normal dan Anomaly .
4. Parameter yang digunakan yaitu 41 fitur pada data NSL KDD 2015.

I.5 Tujuan Penelitian

Adapun tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut :

1. Merancang dan membangun aplikasi klasifikasi serangan jaringan
2. Mengetahui tingkat akurasi dari metode Algoritma *Naïve Bayes* pada klasifikasi jaringan.

I.6 Manfaat Penelitian

Adapun manfaat yang ingin dicapai dari penelitian ini adalah :

1. Membantu mengklasifikasikan jaringan agar dapat diminimalisir serangan jaringan.
2. Membantu penanganan berdasarkan klasifikasi yang dihasilkan sehingga lebih mudah dalam penanganan.
3. Memberi pengetahuan kepada penulis mengenai penggunaan metode Naive Bayes untuk klasifikasi jaringan.

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Dalam menyusun penelitian ini, peneliti menggunakan beberapa penelitian sebelumnya yang ada dalam bentuk jurnal. Jurnal-jurnal yang dipilih tentunya berkaitan serta akan digunakan sebagai perbandingan dengan penelitian yang peneliti lakukan. Jurnal-jurnal yang digunakan antara lain :

Penelitian oleh Bekti Maryuni Susanto (2013) menjelaskan *Naïve Bayes* untuk mendeteksi gangguan jaringan komputer dengan seleksi atribut berbasis korelasi. Hasil yang didapatkan pada penelitian ini diperoleh tingkat akurasi 81,89% dengan seleksi atribut kolerasi (CFS).

Penelitian oleh Trisna, dkk (2015) menjelaskan penerapan *Naïve Bayes* Pada Intrusion Detection Sistem dengan Diskritisasi Variabel. Hasil yang didapatkan pada penelitian ini diperoleh tingkat akurasi 88,20% dengan seleksi atribut kolerasi (CFS).

Penelitian oleh Mohamad Fajarianditya Nugroho, Setyoningsih Wibowo (2017) menjelaskan Fitur Seleksi *Forward Selection* Untuk Menetukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma *Naive Bayes*. Hasil yang didapatkan pada penelitian ini diperoleh tingkat akurasi 97,14% dengan seleksi atribut *Forward Seleksi* untuk klasifikasi status kelulusan mahasiswa dengan atribut yang berpengaruh yaitu status pekerjaan dan IPK semester 4.

Penelitian penulis oleh Sadela (2019) dengan judul Data Mining untuk klasifikasi serangan jaringan menggunakan metode naive bayes. Perbedaan dengan penelitian sebelumnya adalah terdapat pada kasus penelitian yang dilakukan pada CFS (seleksi atribut korelasi) dan pada kasus untuk klasifikasi kelulusan mahasiswa pada UNAKI semarang yang digunakan. Sedangkan persamaanya adalah sama-sama melakukan klasifikasi pada data.

2.2 Landasan Teori

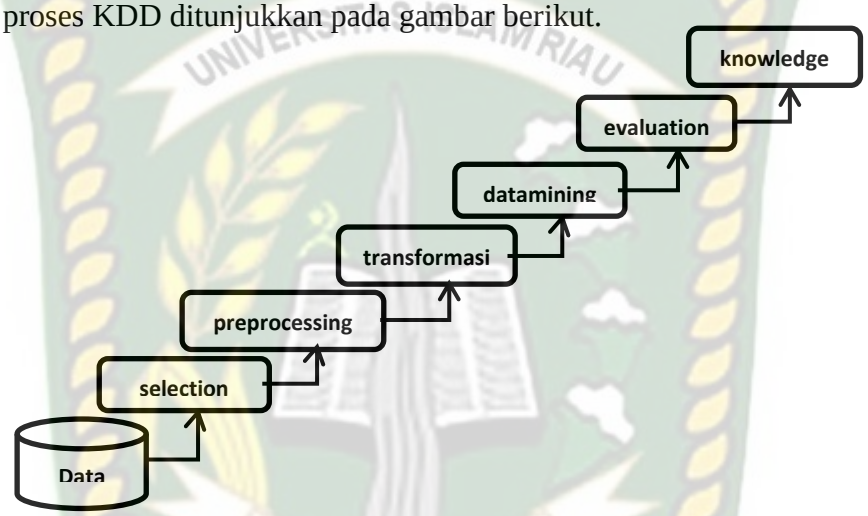
2.2.1 Serangan Jaringan

Keamanan jaringan sangat vital bagi sebuah jaringan komputer. kelemahankelemahan yang terdapat pada jaringan komputer jika tidak dilindungi dan dijaga dengan baik akan menyebabkan kerugian berupa kehilangan data, kerusakan sistem server, tidak maksimal dalam melayani user atau bahkan kehilangan aset-aset berharga institusi.

Serangan yang paling sering digunakan adalah *Port Scanning* dan *DoS (Denial Of Service*) (Ino Anugrah, 2017). *Port Scanning* adalah serangan yang bekerja untuk mencari port yang terbuka pada suatu jaringan komputer, dari hasil *port scanning* akan didapat letak kelemahan sistem jaringan komputer tersebut. *DoS* adalah serangan yang bekerja dengan cara mengirimkan *request* ke server berulang kali untuk bertujuan membuat server menjadi sibuk menanggapi request dan server akan mengalami kerusakan atau hang.

2.2.2 Data Mining

Menurut Han dan Kamber (2007), *Data mining* adalah ilmu yang mempelajari metode untuk menemukan pola dari suatu data. *Data mining* sering juga disebut *knowledge discovery in database* (KDD) yaitu kegiatan pengumpulan data, pemakaina data historis untuk menemukan pola dalam *set* data besar. Adapun proses KDD ditunjukkan pada gambar berikut.



Gambar 2.1 Tahap-tahap Knowledge Discovery from Data (KDD)

Menurut Han (2012) proses dalam KDD secara runtut sebagai berikut :

1. *Data Cleaning*

   Pembersihan data guna menghilangkan *noise* dan data *duplicate.*

2. *Data Integration*

   Proses kombinasi beberapa sumber data. Pada tahap ini dilakukan penggabungan data dari berbagai sumber untuk dibentuk penyimpanan data yang koheren.

3. *Data Selection*

   Proses pengambilan data berkaitan tugas analisis dari basis data.

4. *Data Transformation*

   Pada proses ini data akan diubah kebentuk yang sesuai untuk proses *mining*.

5. *Data Mining*

   Proses ini merupakan proses inti dari KDD, yang merupakan proses pencarian pola atau penggalian informasi dalam data menggunakan teknik tertentu.

6. *Pattern Evaluation*

   Proses ini adalah proses pencocokan pola atau identifikasi kebenaran pola.

7. *Knowledge Presentation*

   Proses representasi secara visual kepada *user* untuk memahami hasil dari *data mining*.

2.2.3 Konsep Klasifikasi

Klasifikasi merupakan metode *data mining* yang digunakan untuk proses pencarian sekumpulan model (fungsi) yang mampu menjelaskan dan membedakan kelas-kelas data atau konsep, yang bertujuan agar model tersebut dapat memprediksi objek kelas yang labelnya tidak diketahui atau dapat memprediksi kecenderungan data-data yang muncul di masa depan. Metode klasifikasi juga bertujuan untuk melakukan pemetaan data ke dalam kelas yang sudah didefinisikan sebelumnya berdasarkan pada nilai atribut data (Han dan Kamber, 2006).

2.2.4 Algoritma Naive Bayes

Naive Bayes Classifier adalah algoritma untuk mencari probabilitas dengan nilai tertinggi untuk pengklasifikasian data uji untuk kategori yang tepat. pada klasifikasi data teks memiliki dua tahapan yaitu data *training* (pelatihan)

kepada dokumen yang sudah diketahui kategorinya, dan tahap kedua yaitu data *testing* (pengujian), adalah proses klasifikasi dokumen yang belum diketahui kategorinya. Pada metode *Naïve Bayes Classifier* data akan dipresesntasikan dengan pasangan atribut "x1, x2…xn" dimana x1 merupakan kata pertama, x2 kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori pada teks. Saat pengklasifikasi maka algoritma akan mencari probabilitas dengan nilai tertinggi pada semua kategori dokumen yang diuji (V MAP) (Jayanti, Sentinuwo, Lantang, & Jacobus, 2016), yang persamaannya adalah :

$$V_{(MAP)} = \underset{V_j eV}{argmax} \frac{\mathrm{P}(x_1,x_2,x_3,\ldots,x_n|V_j)\,P(V_j)}{\mathrm{P}(x_1,x_2,x_3,\ldots,x_n)} \quad (2.1)$$

Pada $\mathrm{P}(x_1, x_2, x_3, \ldots, x_n)$ nilainya konstan disemua kategori (V_j) dengan demikian persamaan dapat ditulis sebagai berikut:

$$V_{(\boldsymbol{MAP})} = \underset{V_j eV}{argmax}\ \mathrm{P}(x_1, x_2, x_3, \ldots, x_n|V_j)\,P(V_j) \quad (2.2)$$

Dan persamaan diatas bisa di Sederhanakan menjadi:

$$V_{(MAP)} = \underset{V_j eV}{argmax}\ \prod_{n=1}^{n} \mathrm{P}(X_i|V_j)\,P(V_j) \quad (2.3)$$

Keterangan:

V_j = Kategori data, j = 1, 2, 3, …, n. Dimana dalam penelitian ini,

J_1 = Kategori Anomali, J_2 = Kategori Normal

$\mathrm{P}(X_i|V_j)$ = Peluang kemunculan X_i pada kategori V_j.

$P(V_j)$ = Peluang kemunculan data yang memiliki kategori j.

Selanjutnya, melakukan perhitungan probabilitas setiap kelas j, dengan persamaan:

$$P(V_j) = \frac{|docs\ j|}{|contoh|} \quad (2.4)$$

Keterangan:

$P(V_j)$: Peluang kemunculan data yang memiliki kategori j.

$|\text{docs}_j|$: Jumlah data perkategori j.

$|\text{contoh}|$: Jumlah data dari semua kategori.

Terakhir melakukan perhitungan probabilitas setiap kata pada data uji terhadap data uji pada setiap kelas j, dengan cara:

$$\text{P}(X_i|V_j) = \frac{n_k+1}{\text{n+ }|\text{jumlah nilai}|} \quad (2.5)$$

Keterangan:

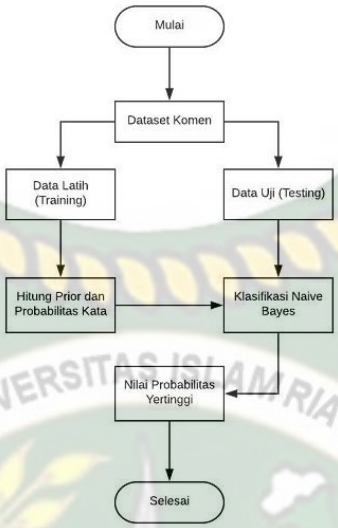
$\text{P}(X_i|V_j)$: Peluang kemunculan X_i pada kategori V_j.

n_k : Jumlah frekuensi kemunculan setiap nilai.

n : Jumlah frekuensi kemunculan nilai dari setiap kriteria.

$|\text{jumlah nilai}|$: Jumlah semua nilai dari semua kriteria.

Dalam tahap ini dilakukan analisa metode *Naïve Bayes*. Dataset dari NSL-KDD diseleksi atribut menggunakan. Tahapan dilakukan mengikuti langkah-langkah dengan algoritma *Naïve Bayes* yaitu:



Gambar 2.2 *Flowchart* Klasifikasi Komentar dengan *Naïve Bayes*

Contoh 13 data dari semua kriteria yang ada sebanyak 41 kriteria/variable yang akan digunakan sebagai data latih untuk menentukan kesimpulan data latih.

Duration	protocol_type	Service	Flag	.....	.....	dst_host_srv_rerror_rate	class
0	tcp	ftp_data	SF	.....	.....	0	normal
0	udp	Other	SF	.....	.....	0	normal
0	tcp	Private	S0	.....	.....	0	anomaly
0	tcp	http	SF	.....	.....	0,01	normal
0	tcp	http	SF	.....	.....	0	normal
0	tcp	Private	REJ	.....	.....	1	anomaly
0	tcp	Private	S0	.....	.....	0	anomaly
0	tcp	Private	S0	.....	.....	0	Anomaly
0	tcp	remote_job	S0	.....	.....	0	anomaly
0	tcp	Private	S0	.....	.....	0	anomaly
0	tcp	Private	REJ	.....	.....	1	anomaly
0	tcp	Private	S0	.....	.....	0	anomaly
0	tcp	http	SF	.....	.....	0	normal

Dalam proses klasifikasi *Naïve Bayes* ini, *dataset* akan dibagi menjadi proses data *training* (latih) dan proses data *testing* (uji). Berikut merupakan penjelasan langkah-langkah dan contoh perhitungan.

1. Data *training* (latih)

Pada tahap ini, nilai pada data yang telah diketahui bobotnya akan dijadikan data latih sebagai acuan dalam membuat model klasifikasi. Kemudian akan dicari nilai probabilitas kategori dan probabilitas masing-masing nilai pada setiap *term* untuk setiap kelas dari data latih. Berikut adalah perhitungannya:

a. Pertama hitung probabilitas setiap kategori (*prior*) seperti pada rumus 2.4, pada penelitian ini yang menjadi kategori ada 2 yaitu kategori NORMAL, dan ANOMALI.

$$\mathrm{P(nor/ano)} = \frac{\mathrm{d(nor/ano)}}{|\mathrm{c}|}$$

$$\mathrm{P(nor)} = \frac{\mathrm{d(nor)}}{|\mathrm{c}|} = \frac{1}{2}$$

$$\mathrm{P(ano)} = \frac{\mathrm{d(ano)}}{|\mathrm{c}|} = \frac{1}{2}$$

b. Kemudian dihitung probabilitas pada setiap *nilai* dari semua data menggunakan rumus persamaan 2.5. jumlah keseluruhan *nilai* yang digunakan pada perhitungan ini sebanyak 13 data (sebagai contoh), 5 *nilai* kelas normal, 8 *nilai* kelas anomali. Berikut adalah merupakan perhitungan probabilitas pada setiap *nilai*:

$$\mathrm{P}(W_k|\mathrm{Nor/Ano}) = \frac{(n_k, \mathrm{Nor/Ano}) + 1}{(\mathrm{n}, \mathrm{Nor/Ano}) + |\mathrm{jumlah\ nilai}|}$$

Diketahui:

|data| = 13

NIlai Normal = 5

NIlai Anomali = 8

Hitung probabilitas dari setiap nilai per kriteria. Mulai dari kriteria *duration* hingga *dst_host_srv_rerror_rate*.

Tabel 2.2 Contoh data training

No.	duration	protocol_type	Service	Flag	.....	.....	dst_host_srv_rerror_rate	class
1.	0	tcp	ftp_data	SF	.....	.....	0	normal
2.	0	udp	Other	SF	.....	.....	0	normal
3.	0	tcp	Private	S0	.....	.....	0	anomaly
4.	0	tcp	http	SF	.....	.....	0,01	normal
5.	0	tcp	http	SF	.....	.....	0	normal
6.	0	tcp	Private	REJ	.....	.....	1	anomaly
7.	0	tcp	Private	S0	.....	.....	0	Anomaly
8.	0	tcp	Private	S0	.....	.....	0	Anomaly
9.	0	tcp	remote_job	S0	.....	.....	0	Anomaly
10.	0	tcp	Private	S0	.....	.....	0	Anomaly
11.	0	tcp	Private	REJ	.....	.....	1	Anomaly
12.	0	tcp	Private	S0	.....	.....	0	Anomaly
13.	0	tcp	http	SF	.....	.....	0	Normal
...	.....	.....	.....	.....	.....	.....	.....	.....
...	.....	.....	.....	.....	.....	.....	.....	.....
...	.....	.....	.....	.....	.....	.....	.....	.....
...	.....	.....	.....	.....	.....	.....	.....	.....
...	.....	.....	.....	.....	.....	.....	.....	.....
65534	0	tcp	Ldap	S0	.....	.....	0	Anomaly

Jumlah Anomaly dari 65534 data = 30713

Jumlah Normal dari 65534 = 34821

a. Probabilitas kriteria "duration"

$$P("0"|"Nor") = \frac{("0"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{64904+1}{34821+65534} = 0{,}64$$

$$P("0"|"Ano") = \frac{("0"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{630+1}{30731+65534} = 0{,}006$$

b. Probabilitas kriteria "protocol type"

$$P("tcp"|"Nor") = \frac{("tcp"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{4+1}{5+13} = \frac{5}{18} = 0{,}27$$

$$P("tcp"|"Ano") = \frac{("tcp"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{8+1}{8+13} = \frac{9}{21} = 0{,}42$$

$$P("udp"|"Nor") = \frac{("udp"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{1+1}{5+1} = \frac{2}{6} = 0{,}33$$

$$P("udp"|"Ano") = \frac{("udp"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{0+1}{8+1} = \frac{1}{9} = 0{,}11$$

c. Probabilitas kriteria "service"

$$P("ftp\_data"|"Nor") = \frac{("ftp\_data"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{1+1}{5+1} = \frac{2}{6} = 0{,}33$$

$$P("ftp\_data"|"Ano") = \frac{("ftp\_data"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{0+1}{8+1} = \frac{1}{9} = 0{,}11$$

$$P("other"|"Nor") = \frac{("other"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{1+1}{5+1} = \frac{2}{6} = 0{,}33$$

$$P("other"|"Ano") = \frac{("other"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{0+1}{8+1} = \frac{1}{9} = 0{,}11$$

$$P("private"|"Nor") = \frac{("private"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{0+1}{5+7} = \frac{1}{12} = 0{,}08$$

$$P("private"|"Ano") = \frac{("private"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{7+1}{8+7} = \frac{8}{15} = 0{,}53$$

$$P("http"|"Nor") = \frac{("http"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{3+1}{5+3} = \frac{2}{8} = 0{,}25$$

$$P(\text{"http"}|\text{"Ano"}) = \frac{(\text{"http"}|\text{"Ano"})+1}{(\text{"Ano"})+|\text{jumlah nilai}|} = \frac{0+1}{8+3} = \frac{1}{11} = 0{,}09$$

$$P(\text{"remote\_job"}|\text{"Nor"}) = \frac{(\text{"remote\_job"}|\text{"Nor"})+1}{(\text{"Nor"})+|\text{jumlah nilai}|} = \frac{0+1}{5+1} = \frac{1}{6} = 0{,}16$$

$$P(\text{"remote\_job"}|\text{"Ano"}) = \frac{(\text{"remote\_job"}|\text{"Ano"})+1}{(\text{"Ano"})+|\text{jumlah nilai}|} = \frac{1+1}{8+1} = \frac{2}{9} = 0{,}22$$

d. Probabilitas kriteria "flag"

$$P(\text{"SF"}|\text{"Nor"}) = \frac{(\text{"SF"}|\text{"Nor"})+1}{(\text{"Nor"})+|\text{jumlah nilai}|} = \frac{5+1}{5+5} = \frac{6}{10} = 0{,}6$$

$$P(\text{"SF"}|\text{"Ano"}) = \frac{(\text{"SF"}|\text{"Ano"})+1}{(\text{"Ano"})+|\text{jumlah nilai}|} = \frac{0+1}{8+5} = \frac{1}{13} = 0{,}07$$

$$P(\text{"SO"}|\text{"Nor"}) = \frac{(\text{"SO"}|\text{"Nor"})+1}{(\text{"Nor"})+|\text{jumlah nilai}|} = \frac{0+1}{5+6} = \frac{1}{11} = 0{,}09$$

$$P(\text{"SO"}|\text{"Ano"}) = \frac{(\text{"SO"}|\text{"Ano"})+1}{(\text{"Ano"})+|\text{jumlah nilai}|} = \frac{6+1}{8+6} = \frac{7}{14} = 0{,}5$$

$$P(\text{"REJ"}|\text{"Nor"}) = \frac{(\text{"REJ"}|\text{"Nor"})+1}{(\text{"Nor"})+|\text{jumlah nilai}|} = \frac{0+1}{5+2} = \frac{1}{7} = 0{,}14$$

$$P(\text{"REJ"}|\text{"Ano"}) = \frac{(\text{"REJ"}|\text{"Ano"})+1}{(\text{"Ano"})+|\text{jumlah nilai}|} = \frac{2+1}{8+2} = \frac{3}{10} = 0{,}3$$

c. Probabilitas kriteria "dst_host_srv_rerror_rate"

$$P(\text{"0"}|\text{"Nor"}) = \frac{(\text{"0"}|\text{"Nor"})+1}{(\text{"Nor"})+|\text{jumlah nilai}|} = \frac{4+1}{5+10} = \frac{5}{15} = 0{,}33$$

$$P(\text{"0"}|\text{"Ano"}) = \frac{(\text{"0"}|\text{"Ano"})+1}{(\text{"Ano"})+|\text{jumlah nilai}|} = \frac{0+1}{8+10} = \frac{1}{18} = 0{,}05$$

$$P(\text{"0.01"}|\text{"Nor"}) = \frac{(\text{"0.01"}|\text{"Nor"})+1}{(\text{"Nor"})+|\text{jumlah nilai}|} = \frac{0+1}{5+1} = \frac{1}{6} = 0{,}16$$

$$P(\text{"0.01"}|\text{"Ano"}) = \frac{(\text{"0.01"}|\text{"Ano"})+1}{(\text{"Ano"})+|\text{jumlah nilai}|} = \frac{1+1}{8+1} = \frac{2}{9} = 0{,}22$$

$$P(\text{"1"}|\text{"Nor"}) = \frac{(\text{"1"}|\text{"Nor"})+1}{(\text{"Nor"})+|\text{jumlah nilai}|} = \frac{0+1}{5+2} = \frac{1}{7} = 0{,}14$$

$$P(\text{"1"}|\text{"Ano"}) = \frac{(\text{"1"}|\text{"Ano"})+1}{(\text{"Ano"})+|\text{jumlah nilai}|} = \frac{2+1}{8+2} = \frac{3}{10} = 0{,}3$$

2. Data *testing* (uji)

Pada proses ini akan dilakukan proses pengujian pada data uji menggunakan data uji pada proses *training* sebelumnya. Pada Tabel 3.2 merupakan contoh data yang menjadi sebagai data uji.

Tabel 2.3 Contoh data testing

duration	protocol_type	Service	Flag	.....	.....	dst_host_srv_rerror_rate	Class
0	tcp	Ldap	S0	.....	.....	0	

a. Probabilitas kriteria "duration"

$$P("0"|"Nor") = \frac{("0"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{0.33+1}{5+13} = \frac{1,33}{18} = 0,07$$

$$P("0"|"Ano") = \frac{("0"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{0.42+1}{8+13} = \frac{1,42}{21} = 0,06$$

b. Probabilitas kriteria "protocol type"

$$P("tcp"|"Nor") = \frac{("tcp"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{0.27+1}{5+13} = \frac{1.27}{18} = 0,07$$

$$P("tcp"|"Ano") = \frac{("tcp"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{0,42+1}{8+13} = \frac{1,42}{21} = 0,06$$

c. Probabilitas kriteria "service"

$$P("idap"|"Nor") = \frac{("idap"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{0+1}{5+13} = \frac{1}{18} = 0,05$$

$$P("idap"|"Ano") = \frac{("idap"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{0+1}{8+13} = \frac{1}{21} = 0,04$$

d. Probabilitas kriteria "flag"

$$P("SO"|"Nor") = \frac{("SO"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{0,09+1}{5+13} = \frac{1,09}{18} = 0,06$$

$$P("SO"|"Ano") = \frac{("SO"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{0,5+1}{8+13} = \frac{1,5}{21} = 0,07$$

e. Probabilitas kriteria "dst_host_srv_rerror_rate"

$$P("0"|"Nor") = \frac{("0"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{0,33+1}{5+13} = \frac{1,33}{18} = 0,07$$

$$P("0"|"Ano") = \frac{("0"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{0,05+1}{8+13} = \frac{1,05}{21} = 0,05$$

Kesimpulan data uji berdasarkan data latih dengan menggunakan rumus persamaan 2.3 yang merupakan rumus hasil penyerderhanaan dari persamaan 2.1 dan 2.2:

V [NORMAL] = 0,07 * 0,07 * 0,05 * 0,06 * 0,07 = 0,0000010

V [ANOMALI] = 0,06 * 0,06 * 0,04 * 0,07 * 0,05 = 0,00000050

Sehingga Kesimpulan dari data uji adalah kategori / class **NORMAL**.

Penjelasan mengenai 41 fitur pada data NSL KDD :

No.	**Kriteria**	**Tipe Data**
1.	Duration	Numerik
2.	protocol_type	Nominal
3.	Service	Nominal
4.	Flag	Nominal
5.	src_bytes	Numerik
6.	dst_bytes	Numerik
7.	Land	Biner
8.	wrong_fragment	Numerik
9.	Urgent	Numerik
10.	Hot	Numerik
11.	num_failed_logins	Numerik
12.	logged_in	Biner
13.	num_compromised	Numerik
14.	root_shell	Numerik
15.	su_attempted	Numerik
16.	num_root	Numerik
17.	num_file_creations	Numerik
18.	num_shells	Numerik

19.	num_access_files	Numerik
20.	num_outbound_cmds	Numerik
21.	is_host_login	Biner
22.	is_guest_login	Biner
23.	Count	Numerik
24.	srv_count	Numerik
25.	serror_rate	Numerik
26.	srv_serror_rate	Numerik
27.	rerror_rate	Numerik
28.	srv_rerror_rate	Numerik
29.	same_srv_rate	Numerik
30.	diff_srv_rate	Numerik
31.	srv_diff_host_rate	Numerik
32.	dst_host_count	Numerik
33.	dst_host_srv_count	Numerik
34.	dst_host_same_srv_rate	Numerik
35.	dst_host_diff_srv_rate	Numerik
36.	dst_host_same_src_port_rate	Numerik
37.	dst_host_srv_diff_host_rate	Numerik
38.	dst_host_serror_rate	Numerik
39.	dst_host_srv_serror_rate	Numerik
40.	dst_host_rerror_rate	Numerik
41.	dst_host_srv_rerror_rate	Numerik

Sumber : Indera Zainul Mutaqien, 2016

2.2.5 *Intrusion Detection System (IDS)*

Intrusion Detection System (IDS) diindikasikan dengan instrumen untuk mendeteksi pelanggaran terhadap kebijakan keamanan sistem. IDS digunakan

untuk mendeteksi adanya trafik paket yang tidak diinginkan pada jaringan. IDS diimplementasikan melalui aplikasi yang terinstal dan mampu memantau paket jaringan untuk mendeteksi paket-paket illegal. (Wu, 2009).

Terdapat banyak jenis IDS yaitu *Network-Based, Wireless IDS, Network Behavior, anomaly Detection* dan *Host-Based.* Pada buku Information *Assurance Tools Report* karangan Wu tahun 2009 ada beberapa jenis pendeteksian pada IDS yaitu *Signature-Based Detection, anomaly-Based Detection,* dan *Stateful Protocol Inspection.*

Signature-Based Detection adalah tipe pendeteksian yang menganalisa potensi terjadinya paket illegal bergantung pada data paket yang diketahui. Tipe pendeteksi ini sangat cepat dan mudah dikonfigurasi. *anomaly-Based Detection* adalah IDS yang memantau paket jaringan dan mendeteksi data yang umumnya tidak normal. Sedangkan *Stateful Protocol Inspection* menyerupai pendeteksi berbasis anomali, tetapi jenis ini dapat menganalisa paket lapisan 3 OSI yaitu lapisan Network dan lapisan 4 yaitu lapisan protokol (Wu, 2009).

Tabel 2.1 Contoh Type Serangan

Type Serangan	KDDTrain	KDDTest
Normal	56252	8421
Dos	35824	732
Probe	11656	2421
U2R	92	210
R2L	985	2754
Total	**104809**	**14538**

2.2.6 PHP *(Hypertext Preprocessor)*

Menurut Agus Saputra dan Feni Agustin (2012:2) PHP memiliki kepanjangan *Hypertext Preprocessor* merupakan suatu bahasa pemograman yang di fungsikan untuk membangun suatu *website* dinamis. PHP di sebut juga sebagai bahasa *Server Side Scripting*. Artinya bahwa dalam setiap menjalankan php, wajib membutuhkan *web server* dan menjalankannya. PHP ini bersifat open source, sehingga dapat di pakai secara gratis dan mampu lintas *platform*, yaitu dapat berjalan pada OS *Windows* maupun *Linux.*

2.2.7 MySQL

MySQL termasuk jenis RDBMS *(Relational Database Management System*). Sehingga istilah seperti tabel, baris dan kolom tetap digunakan dalam MySQL. Pada MySQL sebuah *database* mengandung beberpa tabel, tabel terdiri dari baris dan kolom.

Menurut Sutarman (2007:171) Dalam konteks bahasa SQL, pada umumnya informasi tersimpan dalam tabel-tabel yang secara logika merupakan struktur dua dimensi yang terdiri dari atas baris-baris data (*row* atau *record*) yang berada dalam satu tabel sering disebut sebagai *instance* dari data. Sedangkan kolom sering disebut sebagai *attributes* atau *field*. MySQL mengenal beberapa tipe data, yaitu :

1. Tipe data field
2. Tipe data Numerik
3. Tipe data String

4. Tipe data Tanggal

2.2.8 *XAMPP*

XAMPP merupakan sebuah aplikasi *web server*, juga sering diartikan sebagai layanan data pada *web browser*. Fungsi XAMPP adalah sebagai *server* yang berdiri sendiri (*localhost*), yang terdiri dari beberapa program seperti *Apache* HTTP *Server*, MySQL *database*, dan penerjemah bahasa yang ditulis dengan bahasa pemrograman PHP dan Perl. Program ini tersedia dalam GNU *General Public License* dan bebas, merupakan *web server* yang mudah digunakan yang dapat menampilkan halaman *web* yang dinamis.

2.2.9 *Data Flow Diagram*

Data *Flow Diagram* (DFD) adalah alat pembuatan model yang memungkinkan sistem untuk menggambarkan sistem sebagai suatu jaringan proses fungsional yang dihubungkan satu sama lain dengan alur data, baik secara manual maupun komputerisasi. DFD ini sering disebut juga dengan nama *Bubble chart*, *Bubble diagram*, model proses, diagram alur kerja, atau model fungsi. DFD ini adalah salah satu alat pembuatan model yang sering digunakan, khususnya bila fungsi-fungsi sistem merupakan bagian yang lebih penting dan kompleks dari pada data yang dimanipulasi oleh sistem. Dengan kata lain, DFD adalah alat pembuatan model yang memberikan penekanan hanya pada fungsi sistem.

DFD ini merupakan alat perancangan sistem yang berorientasi pada alur data dengan konsep dekomposisi dapat digunakan untuk penggambaran analisa maupun rancangan sistem yang mudah dikomunikasikan oleh sistem kepada pemakai maupun pembuat program.

Tabel 2.2 Simbol *Simbol Data Flow Diagram*

Simbol	Nama	Fungsi
	Simbol entitas eksternal	Digunakan untuk menunjukkan tempat asal *data* atau *sumber* data.
	Simbol proses	Digunakan untuk menunjukkan tugas atau proses yang dilakukan baik secara manual atau otomatis
	Simbol penyimpanan data	Digunakan untuk menunjukkan Gudang informasi atau data
	Simbol arus data	Digunakan untuk menunjukkan arus dari proses

2.2.10 Program *Flowchart*

Flowchart adalah representasi *grafis* dan langkah-langkah yang harus diikuti dalam menyelesaikan suatu permasalahan yang terdiri dari sekumpulan simbol, dimana masing masing simbol merepresentasikan kegiatan tertentu. *Flowchart* membantu analis dan programmer untuk memecahkan masalah kedalam segmen-segmen yang lebih kecil dan membantu dalam menganalisis alternatif-alternatif dalam pengoprasian.

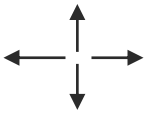
Flowchart diawali dengan penerimaan *input* dan diakhiri dengan penampilan *output. Flowchart* adalah suatu gambaran yang menjelaskan urutan:

1. Pembacaan data.
2. Pemrosesan data.
3. Pengambilan keputusan terhadap data.

4. Penyajian hasil pemrosesan data.

Simbol-simbol *flowchart* yang bisa dipakai adalah simbol-simbol *flowchart standart* yang dikeluarkan oleh *ANSI* dan *ISO*. Berikut ini akan dibahas tentang simbol-simbol yang digunakan untuk menyusun *flowchart* adalah:

Table 2.3 Simbol program *flowchart*

No.	Simbol	Fungsi
1		Terminal, untuk memulai dan mengakhiri suatu proses.
2		Proses, suatu simbol yang menunjukkan setiap pengolahan yang dilakukan oleh computer.
3		*Input-output* untuk memasukkan data atau menunjukkan hasil dari suatu proses.
4		*Decision,* suatu kondisi yang akan menghasilkan beberapa kemungkinan jawaban atau pilihan.
5		*Predefined* proses, suatu simbol untuk menyediakan tempat-tempat pengolahan data dalam *storage.*
6		*Connector,* suatu prosedur akan masuk atau keluar melaui simbil ini dalam lembar yang sama.
7		Arus/*Flow,* prosedur yang dapat dilakukan dari atas kebawah, dari bawah keatas, dari kiri kekanan, dari kanan kekiri.
8		Proses, suatu simbol yang menunjukkan setiap pengolahan yang dilakukan manual.

BAB III

METODOLOGI PENELITIAN

3.1 Alat dan Bahan Penelitian yang Digunakan

Alat dan bahan yang digunakan dalam penelitian ini adalah sebagai berikut:

3.1.1 Alat Penelitian

Pada penelitian ini penulis menggunakan alat dan bahan sebagai pendukung perancangan sistem data mining untuk klasifikasi serangan jaringan dengan metode naive bayes. Adapun kebutuhan spesifikasi perangkat keras untuk perancangan pada penelitian ini adalah :

A. Spesifikasi Kebutuhan *Hardware*

Untuk dapat menjalankan aplikasi dengan baik, tentunya struktur dari perangkat keras (*hardware)* haruslah memenuhi spesifikasi kebutuhan aplikasi yang dibutuhkan, adapun kebutuhan aplikasi terhadap struktur komputer adalah:

1. *Processor* : *Intel Core* i3-4030U
2. Ram : 2,00 GB
3. *Hardisk* : 500 GB
4. *Sysitem Type* : 64-bit *Operating Syatem*

B. Spesifikasi Kebutuhan *Software*

Perangkat lunak (*software*) yang digunakan dalam pembuatan sistem klasifikasi serangan jaringan dengan metode naive bayes adalah :

1. Sistem Operasi : *Microsoft Windows* 8.1 *Pro*
2. Bahasa Pemograman : PHP

3. *Database Management System* (DBMS) : *MySQL*
4. *Web Browser* : *Google Chrome* 61.0
5. Desain Logika Program : *Microsoft Office Visio* 2007

3.1.2 Bahan Penelitian

A. Jenis Data Penelitian

Adapun jenis data yang digunakan dalam penelitian ini adalah data primer yang dikumpulkan melalui wawancara langsung dengan Provider Internet Wanxp, berupa data serangan jaringan yang dipengarui oleh 41 kriteria disamping itu dibutuhkan data sekunder berupa data-data pendukung dalam penelitian ini yang diperoleh dari buku-buku seperti penggunaan metode naive bayes dalam penelitian.

B. Teknik Pengumpulan Data

Adapun teknik pengumpulan data yang diperlukan dalam klasifikasi serangan jaringan dengan metode naive bayes diperoleh dari wawancara dan studi pustaka.

1. Wawancara dilakukan untuk mengumpulkan informasi yang akan berguna dalam mengklasifikasikan serangan jaringan. Wawancara dilakukan pada salah satu karyawan provider internet wanxp.
2. Studi pustaka, mencari referensi-referensi ke pustaka sebagai pedoman penelitian yang penulis lakukan baik berupa buku maupun literatur yang berhubungan dengan penelitian.

3.2 Analisa Sistem yang Sedang Berjalan

Sebelum sistem klasifikasi serangan jaringan dengan metode naive bayes dengan konsep data mining dirancang, sistem yang berjalan masih manual, dalam

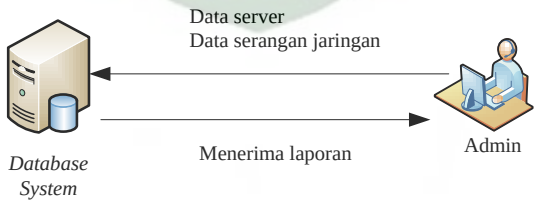
proses pengelompokan serangan dilakukan dengan menganalisa data-data yang masuk ke jaringan wanxp. Admin hanya akan mengkelompokan berdasarkan besarnya jumlah serangan. Analisa sistem yang sedang berjalan bisa dilihat pada gambar 3.1.

Maka dari itu dalam sistem yang sedang berjalan sekarang peneliti selanjutnya akan membuat sistem yang dapat digunakan untuk memudahkan dalam proses klasifikasi serangan jaringan.

Gambar 3.1 Analisa Sitem yang Sedang Berjalan

3.3. Pengembangan Sistem

Dalam penelitian ini akan dirancang sebuah sistem yang akan memudahkan dalam proses klasifikasi serangan jaringan agar dapat di simpulkan untuk penangannya, bisa dilihat pada gambar 3.2.



Gambar 3.2 Pengembangan Sistem

Dari gambar 3.2, dijelaskan bahwa data server dan seranganya di masukan oleh admin yaitu data seluruh server yang di tangani oleh wanxp yang akan

dijadikan kriteria dalam perhitungan klasifikasi serangan jaringan. Selain memasukan data serangan admin juga dapat menambah dan mengurangi member atau user yang dapat menggunakan sistem.

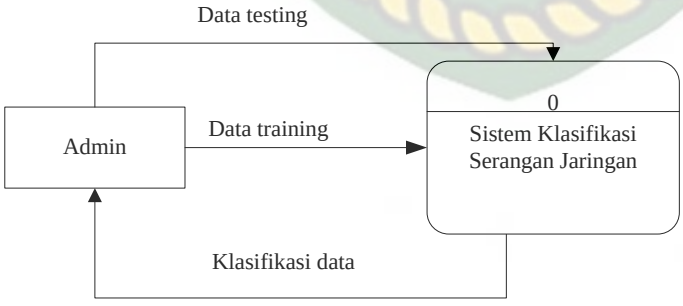
Setelah data masuk kedalam database maka sistem dapat melakukan proses klasifikasi terhadap serangan jaringan. Klasifikasi dilakukan oleh manager agar dapat dapat menghasilkan data klasifikasi serangan jaringan berdasarkan kriteria yang dimasukan dalam sistem.

3.4 Perancangan Sistem

Pada tahap ini akan dijelaskan hal yang berhubungan dengan perancangan sistem yang akan dibuat:

3.4.1 Diagram Konteks

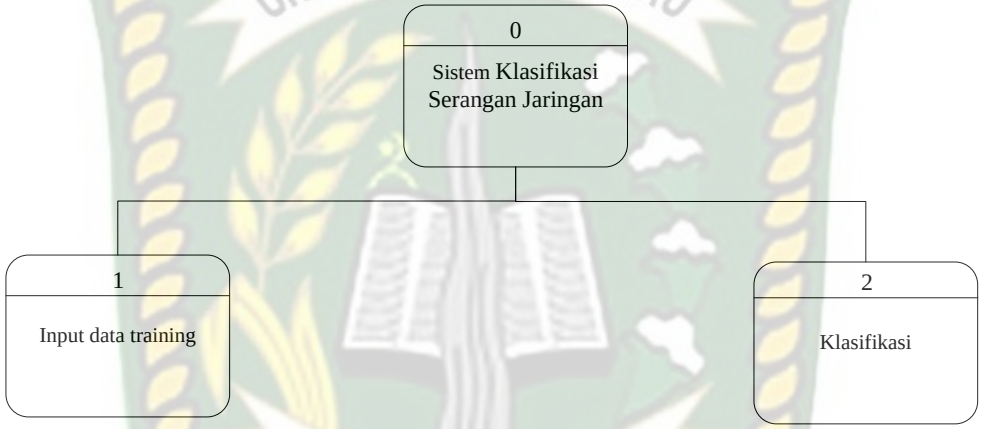
Diagram konteks (*Context Diagram*) digunakan untuk menggambarkan hubungan input dan *output* antara sistem dengan entitas luar, suatu diagram konteks selalu memiliki satu proses yang mewakili seluruh sistem. Sistem ini memiliki dua buah eksternal *entity* yaitu admin dan pengguna.



Gambar 3.3 Diagram Konteks

3.4.2 *Hirarchy Chart*

Hirarchy chart merupakan gambaran subsistem yang menjelaskan proses-proses yang terdapat dalam sistem utama dimana semua subsistem yang berada dalam ruang lingkup sistem utama saling berhubungan satu dan lainnya yang membedakan adalah pada level prosesnya. *Hirarchy chart* sistem yang akan dibangun bisa dilihat pada gambar 3.4.

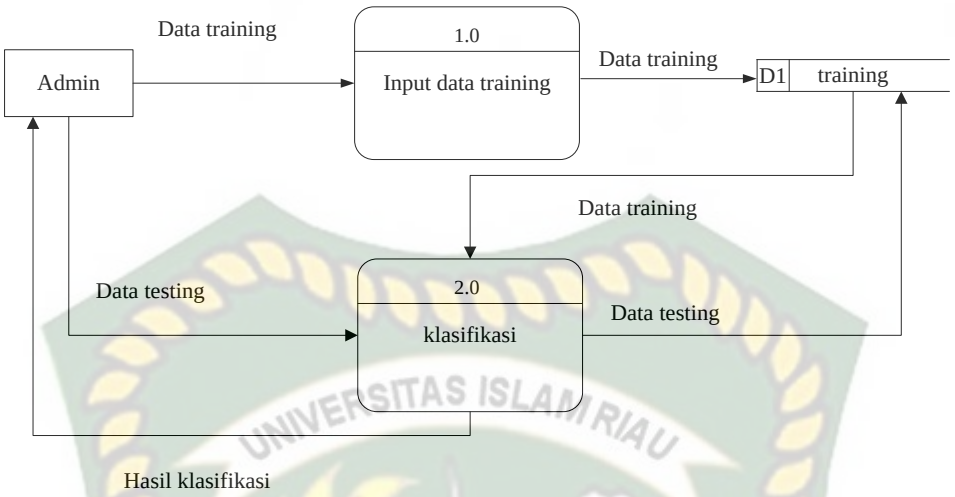


Gambar 3.4 *Hirarchy Chart*

3.4.3 Data *Flow* Diagram (DFD) Level 0

Data *flow* diagram (DFD) akan menjelaskan alur sistem, DFD ini juga akan menggambarkan secara visual bagaimana data tersebut mengalir, pada sistem data mining untuk klasifikasi serangan jaringan ini terdapat beberapa level proses yaitu:

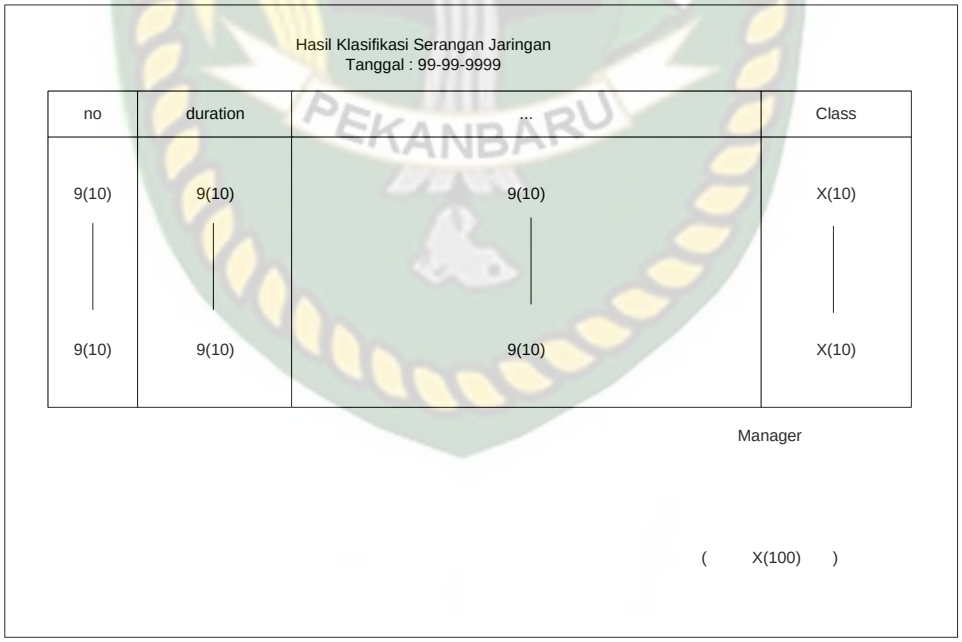
Bisa dilihat pada gambar 3.5 proses pengolahan data terdapat satu penyimpanan data yaitu data training dan data training yang diinputkan oleh admin kemudian disimpan pada *data store*. Sedangkan data hasil klasifikasi akan disaimpan di data training. Selanjutnya dari *data store* data klasifikasi akan diproses oleh sistem dan menghasilkan data klasifikasi serangan jaringan.



Gambar 3.5 DFD Level 0

3.4.4 Desain *Output*

Desain *output* dari data mining untuk klasifikasi serangan jaringan bisa dilihat pada gambar 3.6.



Gambar 3.6 Desain *Output*

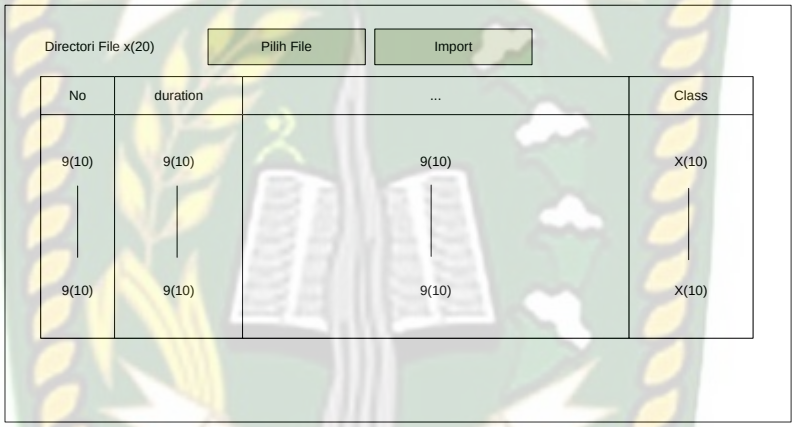
Pada gambar 3.6 hasil sistem klasifikasi serangan jaringan menggunakan data mining dengan metode naive bayes menampilkan tampilan mengenai data

hasil klasifikasi ada 3 hasil klasifikasi serangan jaringan yaitu Bahaya, Darurat, dan Waspada. Sehingga memudahkan dalam penanganan atau penanggulangan serangan jaringan berdasarkan klasifikasinya.

3.4.5 Desain Input

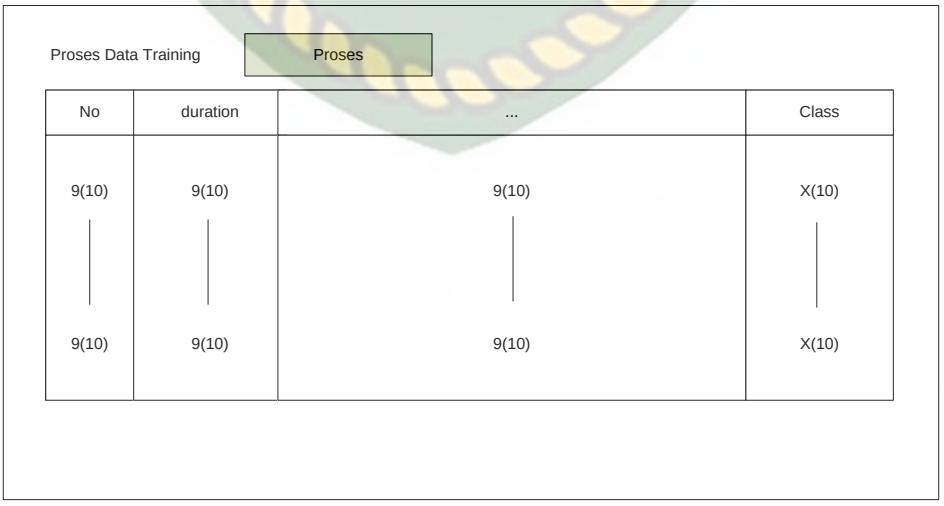
Desain input pada klasifikasi serangan jaringan ini terdiri dari :

1. Desain Input Data Kriteria



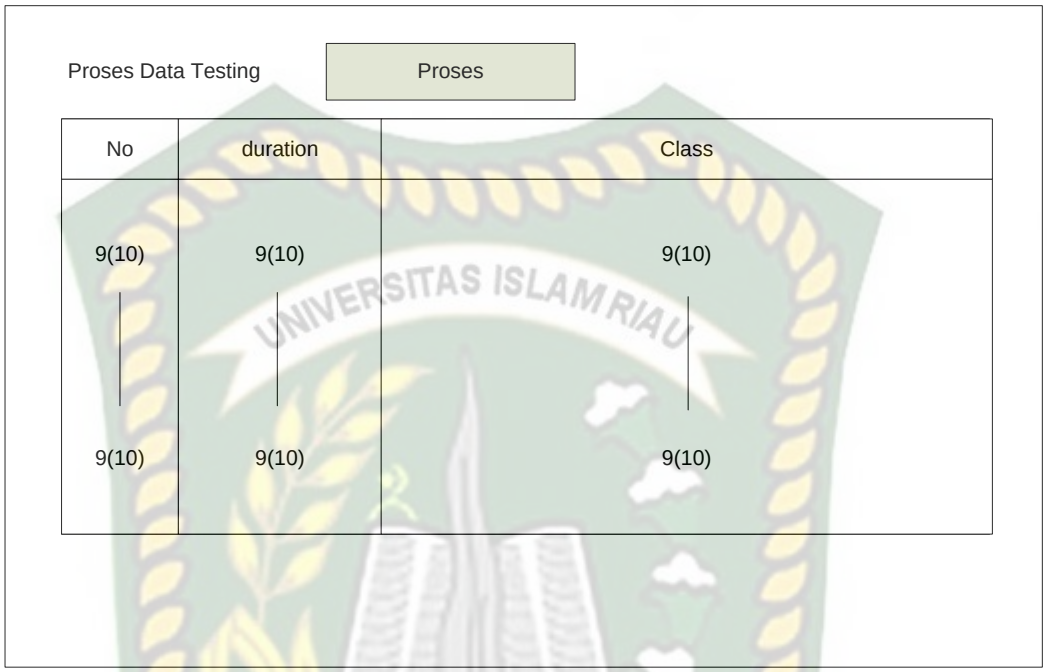
Gambar 3.7 Desain Input kriteria

2. Desain Input Data Training



Gambar 3.8 Desain Input Data Training

3. Rekam Data Testing



Gambar 3.9 Desain Input Data Testing

3.4.6 Perancangan *Database*

3.4.6.1 Skema Data

Pada penelitian ini didapatkan perancangan database dengan nama "db_seranganjaringan" yang terdiri beberapa tabel, antara lain :

1. Tabel Training

   Nama Tabel : training

Tabel 3.1 Desain Tabel training

No	*Field*	*Data Type*	*Size*	Ket
1	Id_training	Int	6	*Primary Key*
2	Duration	Varchar	3	
3	Protocol type	Varchar	2	

4	Service	Varchar	4	
5	Flag	Varchar	10	
6	Srv_byte	Varchar	3	
7	dst_bytes	Varchar	2	
8	Land	Varchar	3	
9	wrong_fragment	Varchar	2	
10	Urgent	Varchar	2	
11	Hot	Varchar	2	
12	num_failed_logins	Varchar	4	
13	logged_in	Varchar	4	
14	num_compromised	Varchar	4	
15	root_shell	Varchar	4	
16	su_attempted	Varchar	4	
17	num_root	Varchar	4	
18	num_file_creations	Varchar	4	
19	num_shells	Varchar	4	
20	num_access_files	Varchar	4	
21	num_outbound_cmds	Varchar	4	
22	is_host_login	Varchar	4	
23	is_guest_login	Varchar	4	
24	Count	Varchar	4	
25	srv_count	Varchar	4	
26	serror_rate	Varchar	4	
27	srv_serror_rate	Varchar	4	
28	rerror_rate	Varchar	4	
29	srv_rerror_rate	Varchar	4	
30	same_srv_rate	Varchar	4	
31	diff_srv_rate	Varchar	4	
32	srv_diff_host_rate	Varchar	4	
33	dst_host_count	Varchar	4	

34	dst_host_srv_count	Varchar	4	
35	dst_host_same_srv_rate	Varchar	4	
36	dst_host_diff_srv_rate	Varchar	4	
37	dst_host_same_src_port_rate	Varchar	4	
38	dst_host_srv_diff_host_rate	Varchar	4	
39	dst_host_serror_rate	Varchar	4	
40	dst_host_srv_serror_rate	Varchar	4	
41	dst_host_rerror_rate	Varchar	4	
42	dst_host_srv_rerror_rate	Varchar	4	
43	Class	Varchar	10	

Contoh data

Tabel Training

Tabel 3.2 Desain Tabel Training

Id_training	Protocol	Service	Flag	Srv_byte	Class
1	Tcp	ftp_data	SF	491	normal
2	Udp	Other	SF	146	normal
3	Tcp	Private	S0	0	anomaly
4	Tcp	http	SF	232	normal
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
65054	Udp	Other	SF	146	normal

BAB IV
HASIL DAN PEMBAHASAN

4.1 Fitur Aplikasi

Fitur aplikasi adalah berbagai menu dan fitur yang telah di sajikan dalam aplikasi yang dapat digunakan oleh pemakai dan menjelaskan tentang kegunaan dari setiap fitur tersebut. Berikut ini fitur dalam aplikasi :

1. Upload Data NSL KDD

   Upload data NSL KDD adalah proses memasukan data dari data NSL dalam format excel. Data yang dimasukan kemudian akan diproses dengan metode Naive Bayes.

2. Proses data training

   Halaman ini adalah untuk memproses data training yang dipilih dari data-data NSL yang sudah di upload sebelumnya. Informasi data training akan membagi data menjadi 70% data training dan 30% data testing.

3. Proses data testing

   Halaman ini adalah untuk memproses data testing yang dipilih dari data-data NSL yang sudah di upload sebelumnya. Informasi data uji akan membagi 30% data testing dari keseluruhan data yang dimasukan.

4. Analisa Data

   Halaman analisa data merupakan halaman proses metode Naive Bayes dilakukan.

   Hasil dari analisa data adalah untuk mengetahui kemiripan data dengan data hasil dan menyimpulkan data normal dan anomaly.

4.2 Pengujian *Black Box*

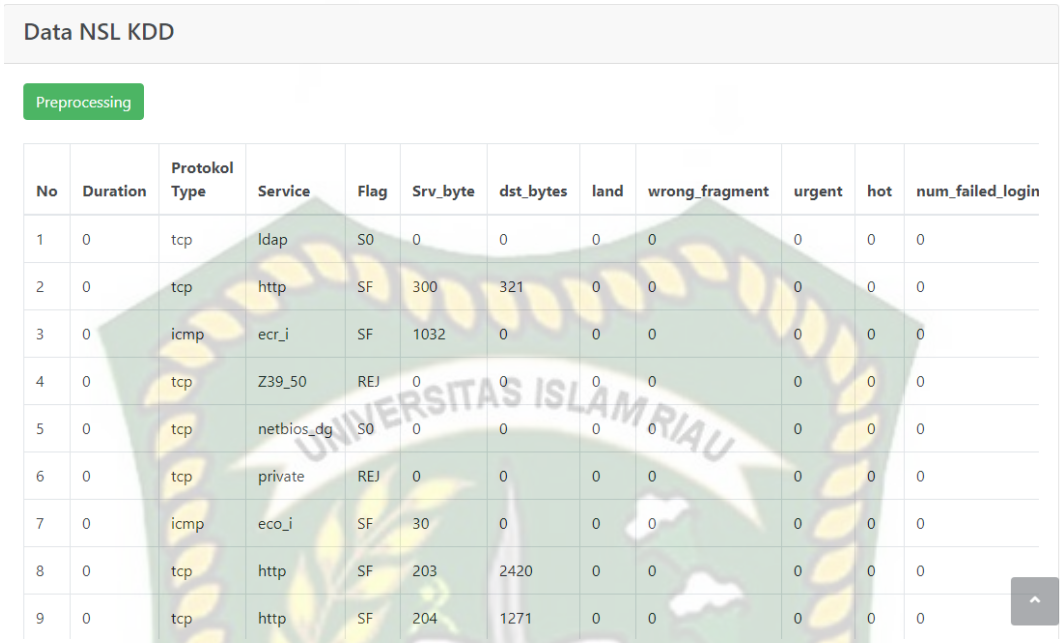
Pengujian *black box (black box testing)* adalah salah satu metode pengujian perangkat lunak yang berfokus pada sisi fungsionalitas, khususnya pada input dan output pada aplikasi untuk menentukan apakah aplikasi tersebut sudah sesuai dengan yang di harapkan.

4.2.1 Pengujian *Form* Upload Data

Pengujian selanjutnya yaitu *form* upload data yang mana dapat dilihat pada gambar 4.1. Pada *form* upload data adalah melihat informasi data NSL yang masuk kedalam sistem.

Gambar 4.1 Pengujian upload data

Pada upload data setelah proses selesai maka hasil dari upload data akan muncul pada menu upload data. Hasilnya sebagai berikut.



Gambar 4.2 Tampilan Hasil Upload Data

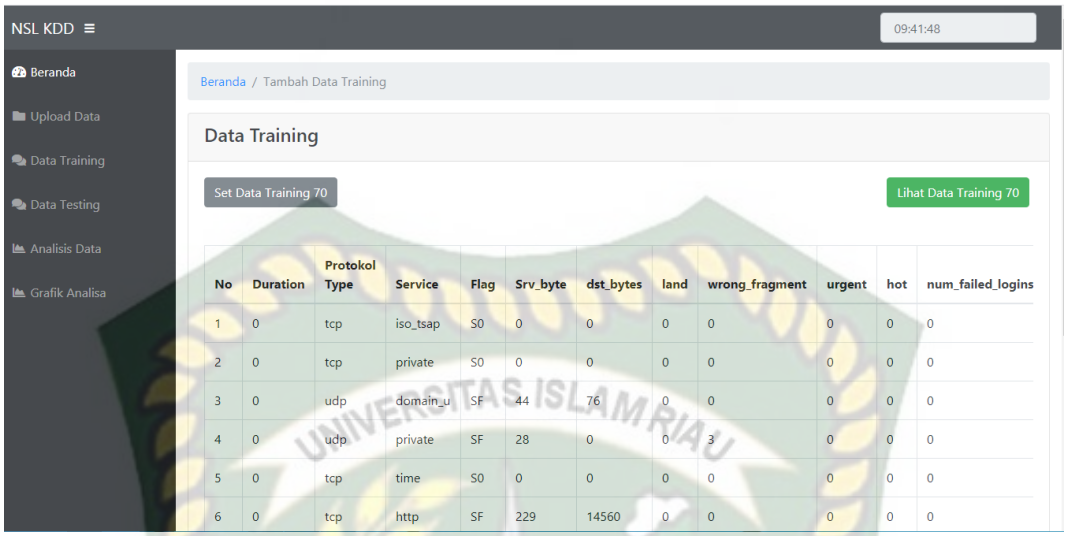
Pada gambar 4.2 adalah hasil proses upload data excel dari data NSL KDD. Pada menu ini kesimpulan dari pengujiannya adalah.

Tabel 4.1 Kesimpulan Pengujian upload data

No.	Komponen yang Diuji	Skenario Pengujian	Hasil yang Diharapkan	Hasil
1	Menampilkan upload data	Memilih menu upload data pada menu	Sistem akan menampilkan upload data	[✓]Sesuai Harapan [ ]Tidak Sesuai Harapan
2	Menampilkan hasil upload data	Menampilkan hasil dari upload data	Sistem akan menampilkan data yang di upload	[✓]Sesuai Harapan [ ]Tidak Sesuai Harapan

4.2.2 Pengujian *Form* Data Training

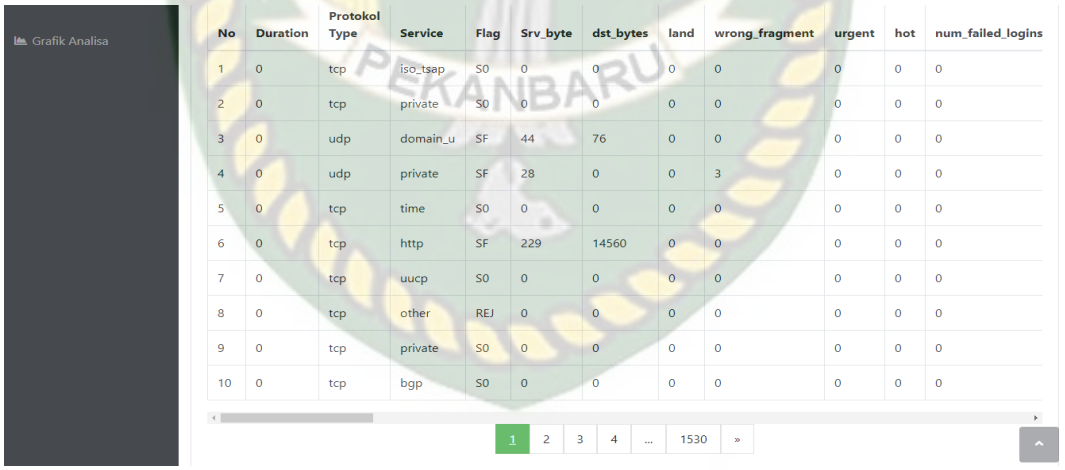
Pengujian selanjutnya yaitu *form* data latih yang mana dapat dilihat pada gambar 4.3. Pada *form* data latih adalah melihat informasi pembagian data menjadi data latih dan data uji. Berikut ini tampilan data latih.



No	Duration	Protokol Type	Service	Flag	Srv_byte	dst_bytes	land	wrong_fragment	urgent	hot	num_failed_logins
1	0	tcp	iso_tsap	S0	0	0	0	0	0	0	0
2	0	tcp	private	S0	0	0	0	0	0	0	0
3	0	udp	domain_u	SF	44	76	0	0	0	0	0
4	0	udp	private	SF	28	0	0	3	0	0	0
5	0	tcp	time	S0	0	0	0	0	0	0	0
6	0	tcp	http	SF	229	14560	0	0	0	0	0

Gambar 4.3 Pengujian data training

Pada data latih klik tombol set data latih setelah proses selesai maka hasil dari pembagian data latih 70% akan muncul pada menu data latih. Hasilnya sebagai berikut.



No	Duration	Protokol Type	Service	Flag	Srv_byte	dst_bytes	land	wrong_fragment	urgent	hot	num_failed_logins
1	0	tcp	iso_tsap	S0	0	0	0	0	0	0	0
2	0	tcp	private	S0	0	0	0	0	0	0	0
3	0	udp	domain_u	SF	44	76	0	0	0	0	0
4	0	udp	private	SF	28	0	0	3	0	0	0
5	0	tcp	time	S0	0	0	0	0	0	0	0
6	0	tcp	http	SF	229	14560	0	0	0	0	0
7	0	tcp	uucp	S0	0	0	0	0	0	0	0
8	0	tcp	other	REJ	0	0	0	0	0	0	0
9	0	tcp	private	S0	0	0	0	0	0	0	0
10	0	tcp	bgp	S0	0	0	0	0	0	0	0

1 2 3 4 ... 1530 »

Gambar 4.4 Tampilan Hasil Data Training

Pada gambar 4.4 adalah hasil proses data training. Pada menu ini kesimpulan dari pengujiannya adalah.

Tabel 4.2 Kesimpulan Pengujian data training

No.	Komponen yang Diuji	Skenario Pengujian	Hasil yang Diharapkan	Hasil
1	Menampilkan data training	Memilih menu data training pada menu	Sistem akan menampilkan data training	[✓]Sesuai Harapan [ ]Tidak Sesuai Harapan
2	Proses data training	Membagi data training 70%	Sistem akan menampilkan data training 70%	[✓]Sesuai Harapan [ ]Tidak Sesuai Harapan

4.2.3 Pengujian *Form* Data Testing

Pengujian selanjutnya yaitu *form* data testing yang mana dapat dilihat pada gambar 4.5. Pada *form* data testing adalah melihat informasi pembagian data menjadi data training dan data testing. Berikut ini tampilan data testing.

Gambar 4.5 Pengujian data testing

Pada data latih klik tombol set data latih setelah proses selesai maka hasil dari pembagian data testing 30% dari keseluruhan data akan muncul pada menu data testing. Hasilnya sebagai berikut.

Gambar 4.6 Tampilan Hasil Data Testing

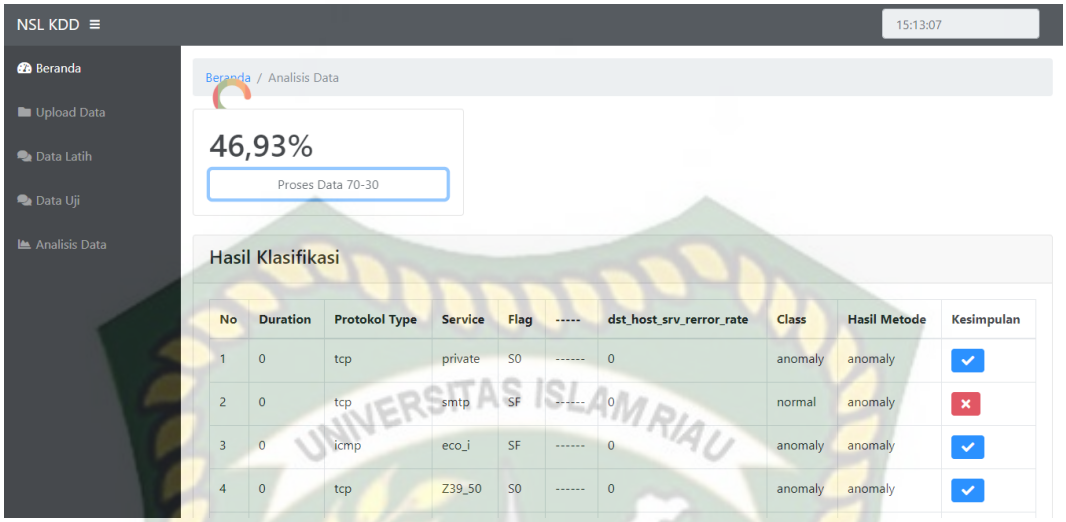
Pada gambar 4.6 adalah hasil proses data testing. Pada menu ini kesimpulan dari pengujiannya adalah.

Tabel 4.3 Kesimpulan Pengujian data testing

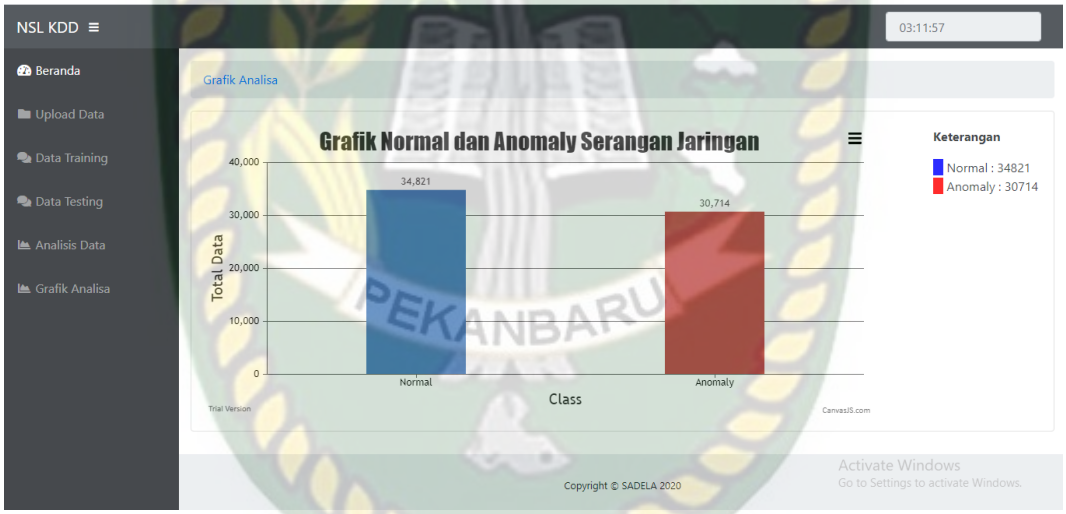
No.	Komponen yang Diuji	Skenario Pengujian	Hasil yang Diharapkan	Hasil
1	Menampilkan data testing	Memilih menu data testing pada menu	Sistem akan menampilkan data testing	[✓]Sesuai Harapan [ ]Tidak Sesuai Harapan
2	Proses data testing	Membagi data testing 30%	Sistem akan menampilkan data testing 30%	[✓]Sesuai Harapan [ ]Tidak Sesuai Harapan

4.2.4 Perbandingan dengan Penggunaan Sistem

Pengujian selanjutnya yaitu *form* perbandingan klasifikasi data yang mana dapat dilihat pada gambar 4.7. Pada *form* perbandingan klasifikasi data adalah memproses metode Naive Bayes untuk melakukan klasifikasi data. Berikut ini tampilan data analisa.



Gambar 4.7 Perbandingan Data Sistem



Gambar 4.8 Grafik analisa

Pada gambar 4.7 adalah hasil proses perbandingan klasifikasi data. Pada proses analisa data kesimpulan data didapatkan dengan berdasarkan proses naive bayes dan dihasilkan data analisa normal dan anomaly pada data hasil. Pada menu ini kesimpulan dari pengujiannya adalah.

Tabel 4.4 Kesimpulan Pengujian analisa data

No.	Komponen yang Diuji	Skenario Pengujian	Hasil yang Diharapkan	Hasil
1	Menampilkan	Memilih menu	Sistem akan	[✓]Sesuai

	analisa data	analisa data pada menu	menampilkan analisa data	Harapan [ ]Tidak Sesuai Harapan
2	Proses analisa data	Menampilkan hasil klasifikasi serangan jaringan	Sistem dapat menampilkan hasil klasifikasi serangan jaringan	[✓]Sesuai Harapan [ ]Tidak Sesuai Harapan

4.2.5 Kesimpulan Pengujian *Black Box*

Dari proses pengujian *black box* ini dapat disimpulkan bahwa setiap data yang akan diinputkan kedalam sistem harus benar-benar sesuai dengan format sistem yang dibuat apabila ada kesalahan dalam penginputan data kedalam sistem, maka sistem akan menolak dan muncul kolom berwarna merah pada *form* yang belum di isi. apabila diinputkan dengan benar sistem dapat berjalan dengan baik dan sesuai dengan harapan yang diinginkan.

4.2.6 Pengujian White Box

Data dari semua kriteria yang ada sebanyak 41 kriteria/variable yang akan digunakan sebagai data latih untuk menentukan kesimpulan data latih. Data yang digunakan berdasarkan kesepakatan untuk pengujian adalah 20 data.

duration	protocol_type	Service	flag	.....	.....	dst_host_srv_r error_rate	class
0	tcp	ftp_data	SF	.....	.....	0	normal
0	udp	Other	SF	.....	.....	0	normal
0	tcp	Private	S0	.....	.....	0	anomaly
0	tcp	http	SF	.....	.....	0,01	normal
0	tcp	http	SF	.....	.....	0	normal
0	tcp	Private	REJ	.....	.....	1	anomaly
0	tcp	Private	S0	.....	.....	0	anomaly
0	tcp	Private	S0	.....	.....	0	anomaly
0	tcp	remote_job	S0	.....	.....	0	anomaly

0	tcp	Private	S0	.....	.....	0	anomaly
0	tcp	Private	REJ	.....	.....	1	anomaly
0	tcp	Private	S0	.....	.....	0	anomaly
0	tcp	http	SF	.....	.....	0	normal
0	tcp	ftp_data	SF	.....	.....	0	anomaly
0	tcp	Name	S0	.....	.....	0	anomaly
0	tcp	netbios_ns	S0	.....	.....	0	anomaly
0	tcp	http	SF	.....	.....	0	normal
0	icmp	eco_i	SF	.....	.....	0	anomaly
0	tcp	http	SF	.....	.....	0	normal
0	tcp	http	SF	.....	.....	0	normal

Dalam proses klasifikasi *Naïve Bayes* ini, *dataset* akan dibagi menjadi proses data *training* (latih) dan proses data *testing* (uji). Berikut merupakan penjelasan langkah-langkah dan contoh perhitungan.

3. Data *training* (latih)

Pada tahap ini, nilai pada data yang telah diketahui bobotnya akan dijadikan data latih sebagai acuan dalam membuat model klasifikasi.

c. Pertama hitung probabilitas setiap kategori (*prior*), pada penelitian ini yang menjadi kategori ada 2 yaitu kategori NORMAL, dan ANOMALI.

$$P(\text{nor/ano}) = \frac{d(\text{nor/ano})}{|c|}$$

$$P(\text{nor}) = \frac{d(\text{nor})}{|c|} = \frac{1}{2}$$

$$P(\text{ano}) = \frac{d(\text{ano})}{|c|} = \frac{1}{2}$$

d. Kemudian dihitung probabilitas pada setiap *nilai* dari semua data menggunakan rumus persamaan 2.5. jumlah keseluruhan *nilai* yang digunakan pada perhitungan ini sebanyak 20 data yang dibagi menjadi

70% data training dan 30% data testing, sehingga didapatkan data training 14 data training dan 6 data testing. Kemudia dicari jumlah *nilai* kelas normal dan jumlah *nilai* kelas anomaly pada data *training* dan *testing*. Berikut adalah merupakan perhitungan probabilitas pada setiap *nilai*:

$$\mathrm{P}(W_k|\mathrm{Nor/Ano}) = \frac{(n_k, \mathrm{Nor/Ano}) + 1}{(\mathrm{n}, \mathrm{Nor/Ano}) + |\mathrm{jumlah\ nilai}|}$$

Diketahui:

|data| = 20

Nilai Normal = 8

Nilai Anomali = 12

Hitung probabilitas dari setiap nilai per kriteria. Mulai dari kriteria *duration* hingga *dst_host_srv_rerror_rate*.

Tabel 4.5 Data training

duration	protocol_type	Service	flag	.....	.....	dst_host_srv_rerror_rate	class
0	tcp	ftp_data	SF	.....	.....	0	normal
0	udp	Other	SF	.....	.....	0	normal
0	tcp	Private	S0	.....	.....	0	anomaly
0	tcp	http	SF	.....	.....	0,01	normal
0	tcp	http	SF	.....	.....	0	normal
0	tcp	Private	REJ	.....	.....	1	anomaly
0	tcp	Private	S0	.....	.....	0	anomaly
0	tcp	Private	S0	.....	.....	0	anomaly
0	tcp	remote_job	S0	.....	.....	0	anomaly
0	tcp	Private	S0	.....	.....	0	anomaly
0	tcp	Private	REJ	.....	.....	1	anomaly
0	tcp	Private	S0	.....	.....	0	anomaly
0	tcp	http	SF	.....	.....	0	normal
0	tcp	ftp_data	SF	.....	.....	0	anomaly

Total data Training = 14

Jumlah Nilai Normal = 5

Jumlah Nilai Anomali = 9

a. Probabilitas kriteria "duration"

$$P("0"|"Nor") = \frac{("0"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{5+1}{5+14} = 0{,}315$$

$$P("0"|"Ano") = \frac{("0"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{9+1}{9+14} = 0{,}4347$$

b. Probabilitas kriteria "protocol type"

$$P("tcp"|"Nor") = \frac{("tcp"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{4+1}{5+14} = 0{,}2631$$

$$P("tcp"|"Ano") = \frac{("tcp"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{9+1}{9+14} = 0{,}4347$$

c. Probabilitas kriteria "service"

$$P("ftp\_data"|"Nor") = \frac{("ftp\_data"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{1+1}{5+14} = 0{,}105$$

$$P("ftp\_data"|"Ano") = \frac{("ftp\_data"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{1+1}{9+14} = 0{,}086$$

d. Probabilitas kriteria "flag"

$$P("SF"|"Nor") = \frac{("SF"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{5+1}{5+14} = 0{,}315$$

$$P("SF"|"Ano") = \frac{("SF"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{1+1}{9+14} = 0{,}086$$

e. Probabilitas kriteria "dst_host_srv_rerror_rate"

$$P("0"|"Nor") = \frac{("0"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{4+1}{5+14} = 0{,}263$$

$$P("0"|"Ano") = \frac{("0"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{7+1}{9+14} = 0{,}347$$

4. Data *testing* (uji)

Pada proses ini akan dilakukan proses pengujian pada data uji menggunakan data uji pada proses *training* sebelumnya.

duration	protocol_type	Service	flag	.....	.....	dst_host_srv_r error_rate	class
0	tcp	Name	S0	.....	.....	0	anomaly
0	tcp	netbios_ns	S0	.....	.....	0	anomaly
0	tcp	http	SF	.....	.....	0	normal
0	icmp	eco_i	SF	.....	.....	0	anomaly
0	tcp	http	SF	.....	.....	0	normal
0	tcp	http	SF	.....	.....	0	normal

Total data Testing = 6

Jumlah Nilai Normal (data training)= 5

Jumlah Nilai Anomali (data training)= 9

a. Probabilitas kriteria "duration"

$$P("0"|"Nor") = \frac{("0"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{3+1}{5+6} = 0{,}363$$

$$P("0"|"Ano") = \frac{("0"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{3+1}{9+6} = 0{,}266$$

b. Probabilitas kriteria "protocol type"

$$P("tcp"|"Nor") = \frac{("tcp"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{3+1}{5+6} = 0{,}363$$

$$P("tcp"|"Ano") = \frac{("tcp"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{2+1}{9+6} = 0{,}2$$

c. Probabilitas kriteria "service"

$$P("name"|"Nor") = \frac{("ftp\_data"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{0+1}{5+6} = 0{,}090$$

$$P("name"|"Ano") = \frac{("ftp\_data"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{1+1}{9+6} = 0{,}133$$

d. Probabilitas kriteria "flag"

$$P("S0"|"Nor") = \frac{("S0"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{0+1}{5+6} = 0{,}090$$

$$P("S0"|"Ano") = \frac{("S0"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{2+1}{9+6} = 0{,}133$$

e. Probabilitas kriteria "dst_host_srv_rerror_rate"

$$P("0"|"Nor") = \frac{("0"|"Nor")+1}{("Nor")+|jumlah\ nilai|} = \frac{3+1}{5+6} = 0{,}363$$

$$P("0"|"Ano") = \frac{("0"|"Ano")+1}{("Ano")+|jumlah\ nilai|} = \frac{3+1}{9+6} = 0{,}2$$

Kesimpulan data uji berdasarkan data latih menggunakan rumus persamaan 2.3 yang merupakan rumus hasil penyerderhanaan dari persamaan 2.1 dan 2.2:

V [NORMAL] = 0,363 * 0,363 * 0,090 * 0,090 * 0,363 = 0,00038

V [ANOMALI] = 0,266 * 0,2 * 0,133 * 0,133 * 0,2 = 0,00018

Sehingga Kesimpulan dari data uji adalah kategori / class **NORMAL**.

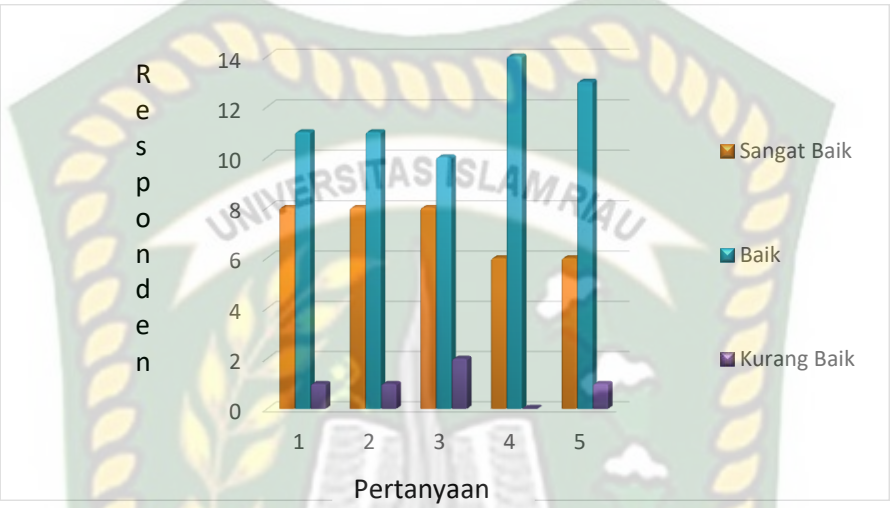
Perhitungan selengkapnya ada di bab Lampiran.

4.3 Implementasi Sistem

Implementasi sistem yang digunakan adalah dengan membuat kuisoner dengan 5 pertanyaan dan 20 responden umum yang terdiri dari pengguna sistem. Kepada 20 responden diajukan pertanyaan-pertanyaan yang dimaksud adalah sebagai berikut :

1. Apakah aplikasi mudah digunakan (*User Friendly*) ?
2. Aplikasi ini mempercepat dan mempermudah dalam proses analisa data serangan jaringan?
3. Bagaimanakah kelengkapan semua fitur dan tampilan aplikasi (*Insert, Delete, dan Layout*) ?
4. Apakah informasi yang diberikan jelas ?
5. Bagaimanakah tingkat keakuratan informasi ?

Dari 5 (lima) pertanyaan diatas, maka diperoleh hasil jawaban atau tanggapan dari responden terhadap kinerja dan tujuan dari sistem pada gambar 4.9.



Gambar 4.9 Grafik Hasil Kuisoner

Keterangan gambar 4.9 :

1. Apakah aplikasi mudah digunakan (*User Friendly*) ? Memiliki nilai Sangat Bagus : 8, Baik : 11, dan Kurang Baik 1.
2. Aplikasi ini mempercepat dan mempermudah dalam proses analisa data serangan jaringan?. Memiliki nilai Sangat Bagus : 8, Baik : 11, dan Kurang Baik 1.
3. Bagaimanakah kelengkapan semua fitur dan tampilan aplikasi (*Insert, Delete, dan Layout*) ? Memiliki nilai Sangat Bagus : 8, Baik : 10, dan Kurang Baik 2.
4. Apakah informasi yang diberikan jelas ? Memiliki nilai Sangat Bagus : 6, Baik : 14, dan Kurang Baik 0.
5. Bagaimanakah tingkat keakuratan informasi ? Memiliki nilai Sangat Bagus : 6, Baik : 13, dan Kurang Baik 1.

4.3.1 Kesimpulan Implementasi Sistem

Berdasarkan hasil kuisoner tersebut maka dapat disimpulkan bahwa sistem ini memiliki nilai sebagai berikut :

Nilai Sangat Baik (SB) akan dikali 5, Nilai Baik (B) akan dikali 3, Nilai Kurang Baik (KB) akan dikali 1 berdasarkan perhitungan skala *likert*.

Tabel 4.5 Hasil Nilai Tiap Pertanyaan Kuisoner

No	Pernyataan	Nilai		
		SB	B	KB
1	Apakah aplikasi mudah digun akan (*User Friendly*) ?	8	11	1
2	Aplikasi ini mempercepat dan mempermudah dalam proses analisa data serangan jaringan?	8	11	1
3	Bagaimanakah kelengkapan semua fitur dan tampilan aplikasi (*Insert, Delete, dan Layout*) ?	8	10	2
4	Apakah informasi yang diberikan jelas ?	6	14	0
5	Bagaimanakah tingkat keakuratan informasi ?	6	13	1

Tabel 4.6 Hasil Perhitungan Tiap Pertanyaan Kuisoner Dengan Skala Likert

Pertanyaan	Nilai		
	SB (x 5)	B (x 3)	KB (x 1)
1	40	33	1
2	40	33	1
3	40	30	2
4	30	42	0
5	30	39	1
Total	180	177	5

Dari hasil kuisioner tabel diatas, yang didasarkan pada 5 pertanyaan yang diajukan secara langsung oleh penulis kepada 20 responden yang diambil secara acak dari pengguna, dapat diambil kesimpulan bahwa aplikasi ini memiliki *performance* baik. Untuk membuktikan akan digunakan menggunakan skala *likert* seperti di bawah ini :

Skor Maksimum = 20 x 5 x 5 = 500 (jumlah responden x skor tertinggi *likert* x jumlah pertanyaan)

Skor Minimum = 20 x 1 x 5 = 100 (jumlah responden x skor terendah *likert* x jumlah pertanyaan)

Indeks (%) = (Total Skor / Skor Maksimum) x 100

Indeks (%) = (362 / 500) x 100

Indeks (%) = 72%

Interval Penilaian

Indeks 0% – 19,99% : Sangat Tidak Setuju

Indeks 20% – 39,99% : Tidak Setuju

Indeks 40% – 59,99% : Ragu-ragu

Indeks 60% – 79,99% : Setuju

Indeks 80% – 100% : Sangat Setuju

Karena nilai Indeks yang kita dapatkan dari perhitungan adalah 72%, maka dapat disimpulkan bahwa responden “SETUJU” sistem dapat dijalankan dengan baik.

BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan hasil analisa klasifikasi data serangan jaringan menggunakan metode naive bayes ini dapat disimpulkan yaitu:

1. Hasil perhitungan manual NSL KDD dan hasil perhitungan program mencapai nilai 82% tingkat kesamaan data antara data dari NSL KDD dan sistem sehingga memiliki performa baik dengan pembagian data training sebanyak 70% yaitu 46873 dan data testing 30% yaitu 18661.
2. Hasil kuisioner di lakukan mendapatkan hasil 72% memberikan hasil pengisian kuisioner adalah Baik dari segi performa dan kualitas dari sistem yang dibangun.

5.2 Saran

Saran dari penulis untuk aplikasi analisa serangan jaringan menggunakan metode naive bayes ini lebih lanjut adalah:

1. Penelitian berikutnya dapat menggunakan metode lain dan teknik penelusuran lain agar aplikasi ini dapat menjadi lebih baik dan dapat melihat hasil perbedaannya.
2. Mengembangkan aplikasi ini agar dapat digunakan lebih mudah dengan berbasis semua *device* atau multiplatform.

DAFTAR PUSTAKA

A.S., Rosa. 2013. "*Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek. Informatika"*. Bandung.

Bekti, M., 2013. "*Naïve Bayes* Untuk Mendeteksi Gangguan Jaringan Komputer dengan Seleksi Atribut Berbasis Kolerasi". Yogyakarta.

Daya, B., 2013. "*Network Security: History, Importance, and Future. "University of Florida Department of Eletrical and Computer Engineering",. http://web.mit.edu/~bdaya/www/Network%20Security*.pdf diakses pada 1 September 2016.

Edi, K., 2016. "Klasifikasi Anomali *Intrusion Detection System* (IDS) Menggunakan *Naïve Bayes* dengan *Wrapper Subset Evaluation* (WRP). Batam".

Fitri. 2009. "*Corporate Governance Concept and Model Preserving True Organization Welfare"*. Yogyakarta: Center for Good Corporate Governance.

Garnaeva, M., dkk. "*Kaspersky Security Bulletin 2016". Statistics,* 2015. *[Online] https://securelist.com/analysis/kaspersky-security-bulletin*/73038/*kasperskysecurity-bulletin*-2015-*overall-statistics-for*-2015/ diakses pada 1 September 2016.

Han, et al. 2007. "*Chapter 11 – Data Mining and Intrusion Detection". Lecture Notes: http://www.cs.uiuc.edu/~hanj/bk2/al3Intrusion*.ppt Diakses tanggal 17 Oktober 2016.

Han, J., et al. 2012. *(3rd Edn). "Data Mining Concepts and Techniques". United States of America: Elsevier Inc.*

Han,J. & Kamber,M. 2006. Data mining: "*Concepts and Techniques", 2nd edition. The Morgan Kaufmann series in Data Management System, Jim Grey, series Editor. http://prdownloads.sourceforge.net/weka/weka-3-8-1jre-x64.exe* diakses pada tanggal 1 Desember 2016. *https://iscxdownloads.cs.unb.ca/iscxdownloads/NSL-KDD/NSL-KDD.zip* diakses pada tanggal 15 September 2016.

Isbat, I., & Moch, Hariadi., 2009. "Pendeteksian Trafik Anomali Pada Jaringan Didasarkan Pada Analisa Payload Data Berbasis Metode *Support Vector Machines"*. Surabaya.

Jabez J & Dr.B.Muthukumar.,2015. "*Intrusion Detection System (IDS): anomaly Detection using Outlier Detection Approach",*. ICCC 48, 338-346.

Jiawei, H., & Micheline, K., , 2006. *"Data Mining: Concepts and Techniques (2nd edition)". Related articles.*

Kadir, Abdul. 2008. "Tuntutan Praktis Belajar *Database* Menggunakan *MySQL*, CV". Andi *Offset.* Yogyakarta.

Maryam, H., 2017. "Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Menggunakan Algoritma *Naïve Bayes* Berbasis *Forward Selection"*. Gorontalo.

Mercury, F., & Adhitya, B., 2018. "Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode *Naïve Bayes* dan *Support Vector Machine* (SVM)". Malang.

Mohamad, F. & Setyoningsih, W., 2017. *"*Fitur Seleksi *Forward Selection* Untuk Menetukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma *Naive Bayes"*. Semarang.

Rian, M., 2012. "Aplikasi Metode Variable Selection Untuk Menentukan Faktor Dominan Yang Mempengaruhi Pendidikan Dan Kesehatan". *http://repository.ipb.ac.id.*

Trisna, dkk., 2015. "Penerapan *Naïve Bayes* Pada *Intrusion Detection System* dengan *Diskritisasi* Variabel". Surabaya.

Wu, Tzeyoung Max. 2009. "Information Assurance Tools Report – Intrusion Detection Systems Sixth Edition". IATAC. Herndon, United States.