

**KLASIFIKASI ANOMALI TRAFIK PADA IDS MENGGUNAKAN
ALGORITMA NAÏVE BAIYES DAN RANDOM FOREST**

PROPOSAL TUGAS AKHIR



Diajukan Oleh :
Asih Asmarani
8020190040

Untuk Persyaratan Penelitian dan Penulisan Tugas Akhir
Sebagai Akhir Proses Studi Strata 1

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS DINAMIKA BANGSA
JAMBI
2022**

IDENTITAS PROPOSAL PENELITIAN

Judul Proposal : Klasifikasi Anomali Trafik Pada IDS
Menggunakan Algoritma Naïve Bayes dan
Random Forest

Program Studi : Teknik Informatika (TI)

Jenjang Pendidikan : Strata 1 (S1)

Peneliti :

- a. Nama Lengkap : Asih Asmarani
- b. NIM : 8020190040
- c. Jenis Kelamin : Perempuan
- d. Tempat/Tgl. Lahir : Jambi / 02 Juli 2000
- e. Alamat : Jln. Darma Karya III Rt.028
Kel.Kenali Asam Bawah
Kec. Kota Baru
- f. No. Telepon : 0895369142465
- g. Email : asihasmarani027@gmail.com

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Pada era sekarang dimana perkembangan teknologi semakin berkembang dengan pesat membuat semakin meningkatnya segala aktifitas yang erat kaitannya akan jaringan, baik kebutuhan akses informasi ataupun data. Penggunaan jaringan yang semakin besar dengan kurangnya upaya menjaga keamanan jaringan, akan membuka peluang terjadinya suatu ancaman, serangan bahkan tindakan peretasan jaringan (*hacker*). Keamanan jaringan menjadi suatu yang penting dimana seiring dengan kebutuhan akan informasi yang terdapat pada jaringan dengan melakukan upaya pengamanan membuat data yang ada menjadi aman, salah satu upaya yang dapat dilakukan yaitu melakukan deteksi terhadap suatu serangan yang merupakan awal dari serangan terhadap sistem yang dilakukan oleh peretas.[1]

Berdasarkan [2] dan [3] Penerapan IDS (*Intrusion Detection System*) merupakan salah satu upaya menjaga keamanan sistem jaringan serta tujuannya yaitu menentukan atau mengidentifikasi kapan anomali terjadi dan asal anomali, sehingga membantu mengurangi dampak dari serangan. Jenis serangan yang dapat mengancam suatu keamanan sistem banyak jenisnya diantaranya yaitu serangan *DoS*, *DdoS*, *Brute Force*, *Sql Injection*, dll. IDS dapat memberikan laporan mengenai tipe-tipe serangan

yang sedang terjadi yang disimpan dalam sebuah log sehingga celah pada sistem dapat diperbaiki dan diperkuat sistem keamanannya.

Dari data-data yang diperoleh pada data serangan *Intrusion Detection System* dapat diolah untuk didapatkan suatu informasi atau pengetahuan. *Data Mining* biasa dimanfaatkan dalam pengolahan data, untuk mengetahui pola yang penting atau menarik dari data yang ada di *database* ataupun *dataset* yang besar.[4] Metode klasifikasi biasa digunakan dalam pengidentifikasian suatu pola data, *Naïve Bayes* dan *Random Forest* merupakan salah satu dari algoritma klasifikasi. Berdasarkan latar belakang yang telah dijabarkan sebelumnya, melatar belakangi penulis untuk melakukan penelitian dengan judul “**KLASIFIKASI ANOMALI TRAFIK PADA IDS MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN RANDOM FOREST**”.

1.2 RUMUSAN MASALAH

Adapun rumusan masalah yang akan diselesaikan dalam penelitian ini, sebagai berikut :

- a. Bagaimana mendeteksi trafik anomali pada trafik jaringan.
- b. Bagaimana menerapkan algoritma *Naïve Bayes* dan *Random Forest* dalam mengenali trafik anomali pada trafik jaringan.
- c. Bagaimana performa klasifikasi algoritma *Naïve Bayes* dan *Random Forest* dalam mendeteksi trafik anomali pada trafik jaringan.

1.3 BATASAN MASALAH

Batasan masalah yang dilakukan pada penelitian ini yaitu :

- a. *Dataset* yang digunakan yaitu menggunakan *dataset CICIDS2017*.
- b. Serangan yang dideteksi yaitu trafik anomali *Web Attack Brute Force, SQL Injection* dan *XSS* yang ada pada trafik jaringan.
- c. Menggunakan algoritma *Naïve Bayes* dan *Random Forest* dalam mengklasifikasi trafik anomali.
- d. Seleksi fitur (*Feature Selection*) yang digunakan pada penelitian ini yaitu *Information Gain*.

1.4 TUJUAN PENELITIAN

Berikut tujuan yang dapat dicapai dari dilakukan penelitian ini yaitu :

- a. Membedakan data trafik normal dan data serangan (trafik anomali) pada trafik jaringan.
- b. Menghitung akurasi deteksi serangan pada trafik jaringan.
- c. Membandingkan performa algoritma *Naïve Bayes* dan *Random Forest* dalam mendeteksi serangan atau trafik anomali pada trafik jaringan.

1.5 MANFAAT PENELITIAN

Adapun manfaat yang didapatkan dari penelitian ini yaitu :

- a. Menghasilkan metode deteksi serangan atau trafik anomali dengan performa pengklasifikasian yang baik.
- b. Sebagai rujukan bagi peneliti di bidang *Intrusion Detection System*.

1.6 SISTEMATIKA PENULISAN

Untuk memberikan suatu gambaran yang jelas mengenai keseluruhan penulisan ilmiah dibuatlah sistematika penulisan, agar dalam penyusunan penelitian dapat dilakukan secara sistematis membahas topik yang di angkat dan menghindari terjadinya pembahasan diluar judul penelitian, dapat dilihat sistematika penulisan berikut meliputi :

BAB I : PENDAHULUAN

Bab ini membahas latar belakang, perumusan masalah, pembatasan masalah, tujuan dan manfaat penelitian serta sistematika penulisan.

BAB II : LANDASAN TEORI

Bab ini berisikan teori-teori yang mendukung penelitian dan berisikan definisi yang melandasi penelitian yang didapatkan dari studi literatur baik dikutip dari buku, jurnal dan lain-lain, serta berisi tinjauan penelitian sejenis.

BAB III : METODOLOGI PENELITIAN

Bab ini berisikan tentang kerangka kerja penelitian, metode yang digunakan, rancangan eksperimen yang akan dilakukan serta *tools* yang akan dipakai dalam penelitian sebagai alat bantu dalam menjawab permasalahan penelitian.

BAB IV : ANALISIS DAN PEMBAHASAN

Bab ini berisikan tentang analisa terhadap *dataset* yang dipakai atau terhadap *dataset* CICIDS2017 meliputi analisa *dataset*, analisa trafik jaringan, deteksi serangan menggunakan metode yang dipilih, dan berisi hasil analisa dari uji coba yang telah dilakukan atau hasil deteksi anomali trafik (serangan) yang ada pada *dataset* CICIDS2017.

BAB V : PENUTUP

Bab ini berisikan kesimpulan dan saran dari penelitian yang telah dilakukan untuk pengembangan penelitian lebih lanjut.

BAB II

LANDASAN TEORI

2.1 ANOMALI TRAFIK

Menurut [5] dan [6] Anomali trafik merupakan suatu keadaan yang menyebabkan abnormalisasi pada lalu lintas jaringan. Penyebab dari anomali ini bisa saja faktor dari banyaknya pengguna internet atau serangan pada suatu jaringan atau adanya aktivitas-aktivitas dalam jaringan yang menyimpang dari batas normal. Kondisi ini dapat menyebabkan penurunan performansi jaringan sehingga rentannya sebuah jaringan untuk diserang, sehingga perlu dilakukan deteksi anomali yang terjadi.

2.2 INTRUSION DETECTION SYSTEM (IDS)

Niko dkk. [7] menyatakan *Intrusion Detection System* merupakan proses memonitor trafik jaringan dalam sebuah sistem untuk mendeteksi adanya pola dan aktivitas yang mencurigakan yang memungkinkan adanya serangan dalam suatu sistem tersebut.

Berdasarkan [8] *Intrusion detection system* adalah suatu perangkat lunak (*software*) atau suatu sistem perangkat keras (*hardware*) yang bekerja secara otomatis untuk memonitor kejadian pada jaringan komputer dan dapat menganalisis keamanan jaringan. IDS memiliki 3 (tiga) komponen fungsi fundamental yang merupakan proses utama dalam IDS. Komponen fungsi tersebut yaitu :

1. Pengambilan Data (*Information Sources*). Komponen ini merupakan fungsi untuk melakukan pengambilan data dari berbagai sumber yang ada pada sistem yang diamati.
2. Analisis. Bagian ini melakukan organisasi terhadap data yang diperoleh, mengambil kesimpulan terhadap pelanggaran / *intrusion* baik yang sedang terjadi maupun yang telah terjadi.
3. Respon. Komponen ini melakukan beberapa aksi pada sistem setelah pelanggaran yang terjadi telah terdeteksi. Respon ini dapat dikelompokkan menjadi 2 (dua) yaitu respon aktif dan respon pasif. Respon aktif berupa melakukan beberapa aksi secara otomatis untuk mengintervensi sistem yang ada, sedangkan respon pasif memberikan *report* pada administrator yang akan melakukan respon terhadap sistem.

2.2.1 Kategori IDS

Berdasarkan [7], [9] dan [10] IDS diklasifikasikan dalam 2 kategori teknik deteksi intrusi berdasarkan bagaimana data dianalisis, yaitu :

a. Misuse Detection

Sistem mempelajari pola penyerangan yang ada dan sudah dikenal. Pola ini dipelajari dengan memeriksa seluruh data yang datang untuk menemukan tipe *intrusion*. Metode ini tidak mampu mengidentifikasi tipe serangan baru yang polanya belum diketahui.

b. Anomaly Detection

Menggunakan pendekatan terhadap pola data normal untuk mendeteksi instruksi. Pola dipelajari dari data normal, dimana data yang tidak terlihat dicek dan dicari penyimpangan dari pola yang telah dipelajari. Jika data yang terekam menyimpang dari data normal maka dianggap sebagai serangan.

2.3 Jenis Deteksi Anomali

Berdasarkan [9] dan [10] ada dua jenis IDS jika dilihat kemampuan mendeteksi serangan di dalam jaringan yaitu *host-based* IDS (HIDS), dan *network-based* IDS (NIDS).

1. *Host-Based IDS (HIDS)*

Merupakan *system* yang mampu mendeteksi hanya pada *host* tempat implementasi *intrusion detection system*. Aktivitas sebuah *host* jaringan individu akan dipantau apakah terjadi percobaan serangan atau penyusupan ke dalamnya atau tidak. HIDS ditempatkan pada sebuah *device* seperti *server* atau *workstation*, dimana data yang dianalisa berada pada mesin lokal.

2. *Network-Based (NIDS)*

Merupakan *system* yang akan menganalisis semua lalu lintas yang melewati ke sebuah jaringan yang akan mencari adanya percobaan serangan atau penyusupan ke dalam *system* jaringan, atau dengan kata lain NIDS menganalisa semua trafik data dalam sebuah jaringan dalam mendeteksi serangan. *Intrusion detection system* menggunakan adapter *promiscuous mode*

sehingga dapat melihat dan menganalisa semua trafik paket yang melewati jaringan secara *realtime*.

2.4 DATASET CICIDS2017

Berdasarkan [11] dan [12] CICIDS2017 merupakan kumpulan *dataset* yang berisikan serangan umum yang jinak dan paling mutakhir, *dataset* ini menyerupai data sebenarnya di dunia nyata yang berisikan hasil analisis lalu lintas jaringan yang terdiri dari trafik normal dan trafik anomali atau data serangan.

Tabel 2.1 Detail Data Dataset CICIDS2017

Nama File	Jenis Trafik	Jumlah Record
Monday-WorkingHours.pcap_ISCX.csv	Benign	529,918
Tuesday-WorkingHours.pcap_ISCX.csv	Benign	432,074
	SSH-Patator	5,897
	FTP-Patator	7,938
Wednesday-WorkingHours.pcap_ISCX.csv	Benign	440,031
	DoS Hulk	231,073
	DoS GoldenEye	10,293
	DoS Slowloris	5,796
	DoS Slowhttptest	5,499
	Heartbleed	11
Thursday-WorkingHours-Morning-WebAttacks.pcap_ISCX.csv	Benign	168,186
	WebAttack-Brute Force	1,507
	WebAttack-Sql	21
		652

	Injection Web Attack-XSS	
Thursday-WorkingHours- Afternoon- Infiltration.pcap_ISCX.csv	Benign Infiltration	288,566 36
Friday-WorkingHours- Morning.pcap_ISCX.csv	Benign Bot	189,067 1,966
Friday-WorkingHours- Afternoon- PortScan.pcap_ISCX.csv	Benign PortScan	127,537 158,930
Friday-WorkingHours- Afternoon-DDos pcap_ISCX.csv	Benign DdoS	97,718 128,027
Total Instance/Record		2.830,743

2.5 DATA MINING

Data mining didefinisikan sebagai satu set teknik yang digunakan secara otomatis untuk mengeksplorasi secara menyeluruh dan membawa ke permukaan relasi-relasi kompleks pada set data yang sangat besar, atau secara sederhana data mining mengekstrasi informasi atau pola yang penting atau menarik dari data yang ada di *database* yang besar dan biasa dikenal dengan nama *Knowledge Discovery in Database (KDD)*. [4]

2.5.1 Metode Data Mining

Berdasarkan [13] terdapat beberapa teknik yang digunakan dalam *data mining* diantaranya yaitu :

a. Prediksi

Pada *data mining* untuk prediksi, pemodelan dilakukan menggunakan data sampel yang diketahui nilai atributnya untuk memperkirakan nilai atribut dari target tertentu. Dengan adanya model maka dapat dibuat aplikasi yang sesuai untuk setiap kondisi.

b. Klasifikasi

Klasifikasi adalah penggalian data yang menetapkan *item* dalam kolerasi untuk menargetkan kelas tertentu. Klasifikasi dimulai dengan mengumpulkan data dimana data tersebut terdapat kelas yang sudah diketahui.

c. Regresi

Regresi mirip dengan klasifikasi, perbedaannya yaitu regresi tidak bisa mencari pola yang dijelaskan dalam bentuk kelas. Regresi juga dapat digunakan untuk prediksi. Tujuan dari regresi adalah mencari pola dan menentukan suatu nilai numerik.

d. *Clustering*

Clustering digunakan untuk menemukan kelompok objek data yang serupa. Kelompok objek *cluster* terdiri dari anggota *cluster* yang lebih mirip satu sama lain daripada anggota kelompok *cluster* lainnya. Data anggota kelompok

dibagi ke dalam kelompok-kelompok yang belum ditentukan. Data yang digunakan untuk *clustering* tidak mempunyai kelas/target. *Clustering* berfungsi sebagai *preprocessing*, dimana *clustering* untuk mengidentifikasi anggota kelompok yang mirip, yang dapat digunakan untuk membangun model yang diawasi.

e. Asosiasi

Asosiasi digunakan untuk menemukan probabilitas antara item dengan kumpulan item. Hubungan antara item dinyatakan sebagai aturan asosiasi yang dapat digunakan untuk menentukan atau menganalisis pola pada serangkaian kejadian. Pemodelan asosiasi biasanya disebut *market-basket analysis*.

2.5.2 Tahapan Data Mining

Berdasarkan [14] *Data mining* memiliki tahapan-tahapan untuk mencari pengetahuan terhadap suatu data atau disebut *Knowledge Discovery In Databases (KDD)*. KDD merupakan proses untuk mengetahui pola kumpulan data dengan jumlah besar. Secara umum tahapannya terdiri dari :

- a. *Data Cleaning*. Proses menghilangkan *noise* dari data yang tidak konsisten.
- b. *Data Integration*. Penggabungan data dari berbagai *database* ke dalam satu *database* baru.
- c. *Data Selection*. Proses pemilihan data yang relevan yang didapat dari *database*.

- d. *Data Transformation*. Data diubah ke dalam format yang sesuai untuk diproses dalam *data mining*.
- e. *Data Mining*. Suatu metode yang diterapkan untuk menemukan pengetahuan berharga yang tersembunyi dari data.
- f. *Pattern Evaluation*. Mengidentifikasi pola-pola menarik untuk dipresentasikan ke dalam *knowledge based*.
- g. *Knowledge Presentation*. Visualisasi dan penyajian pengetahuan mengenai teknik yang digunakan untuk memperoleh pengetahuan yang diperoleh oleh *user*.

2.6 KLASIFIKASI

Berdasarkan [15] Klasifikasi data terdiri dari dua langkah proses, yang pertama adalah proses *learning (fase training)* dimana algoritma klasifikasi dibuat untuk menganalisa data training lalu direpresentasikan dalam bentuk *rule* klasifikasi, proses kedua adalah klasifikasi dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi. Proses klasifikasi didasarkan pada empat komponen yaitu [16] :

- a. Kelas yaitu variabel dependen yang berupa kategorikal yang merepresentasikan label yang terdapat pada objek.
- b. *Predictor* yaitu variabel independen yang direpresentasikan oleh karakteristik atribut data.

- c. *Training dataset* yaitu satu set data yang berisi nilai dari kedua komponen diatas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.
- d. *Testing dataset* yaitu data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

2.7 ALGORITMA NAÏVE BAYES

Karakteristik *Naïve Bayes* dalam melakukan klasifikasi didasarkan pada teori probabilitas yang memandang semua fitur atau atribut dari data sebagai bukti probabilitas, kaitan antara *Naïve Bayes* dengan klasifikasi, korelasi hipotesis dan bukti dengan klasifikasi adalah bahwa teorema Bayes merupakan label kelas yang menjadi target pemetaan [17].

Berdasarkan [7] *Naïve Bayes* mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah idependen secara bersyarat yang diberikan dengan label kelas. Tahapan algoritma *naïve bayes* adalah sebagai berikut :

1. Menyiapkan data *training*.
2. Setiap data dipresentasikan sebagai vektor berdimensi-n yaitu $X = X_1, X_2, X_3$
..... X_n
3. N adalah gambaran dari ukuran yang dibuat di test dari n atribut yaitu $A_1, A_2,$
 A_3 A_n
4. M adalah kumpulan kategori yaitu $X = C_1, C_2, C_3$ C_m

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 C_1 jika dan hanya jika $P(C_i|X) > P(C_j|X)$ untuk $1 \leq j \leq m, j \neq i$
7. Kemudian kita perlu memaksimalkan $P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)}$.
8. Dimana x adalah nilai-nilai atribut dalam sampel X dan probabilitas $P(x_1|C_i), P(x_2|C_i), \dots, P(x_n|C_i)$, dapat diperkirakan dari *data training*

2.8 ALGORITMA RANDOM FOREST

Berdasarkan [18] dan [19] *Random Forest* (hutan acak) adalah sekumpulan pengklasifikasi yang terdiri dari beberapa pohon keputusan dan diklasifikasikan berdasarkan hasil klasifikasi untuk setiap anggota pohon keputusan. Metode hutan acak merupakan perpanjangan dari metode CART, yang terdiri dari penggunaan agregasi *bootstrap* (*bagging*) dan metode pemilihan fitur secara acak. Beberapa pohon tumbuh di hutan acak untuk membentuk hutan, kemudian dilakukan analisis terhadap kelompok pohon tersebut. Dalam kumpulan data yang terdiri dari n variabel observasi dan penjelas, hutan acak diimplementasikan oleh, (1) lakukan n pengambilan sampel acak dengan pemulihan set data. Tahap ini adalah tahap *bootstrap*, (2) dalam contoh *bootstrap*, pohon dibuat hingga mencapai ukuran maksimumnya (tanpa pemangkasan). Pada setiap node, penyortiran dilakukan dengan memilih m variabel penjelas secara acak, di mana $m \ll p$. Penyortir terbaik dipilih

dari variabel penjelas ini. Langkah ini adalah langkah pemilihan fitur acak, dan (3) ulangi langkah 1 dan 2 sampai k kali untuk membuat hutan dengan k pohon.

2.9 WEKA

Menurut Faid [18] Weka merupakan rangkain perangkat lunak pembelajaran mesin yang ditulis dalam bahasa *Java* dan dikembangkan di Universitas Waikato, Selandia Baru. Perangkat lunak ini memiliki banyak algoritma *machine learning* untuk keperluan *data mining*. Weka juga memiliki banyak *tools* untuk pengolahan data, mulai dari *preprocessing*, *classification*, *association rules*, dan *visualization*.

2.10 PENELITIAN SEJENIS

Berikut merupakan penelitian sejenis yang dilakukan sebelumnya oleh peneliti-peneliti lain dan digunakan sebagai sumber acuan atau referensi untuk membuat penelitian ini, diantaranya yaitu :

No	Peneliti/Tahun	Judul Penelitian	Metode	Hasil
1	Niko Suwaryo, Ismasari Nawangsih, Sri Rejeki / 2021 [7]	Deteksi Serangan Pada Intrusion Detection System (Ids) Untuk Klasifikasi Serangan Dengan Algoritma Naïve Bayes, C.45 Dan K-Nn Dalam	Algoritma Naïve Bayes, C.45 Dan K-Nn	Hasil pegujian dari algoritma Naïve Bayes, K-Nearest Neighbor dan C.45, C.45 mendapatkan hasil lebih baik

		Meminimalisasi Resiko Terhadap Pengguna		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 seharusnya No (Tidak ada serangan), sedangkan
--	--	---	--	--

				<p>algoritma C,45, atribut (attack) normal, dos, probe dan r2l, normal (tidak ada serangan), yes (adanya serangan) sehingga menghasilkan tingkat accuracy 97.80%, recall 98.18% dan 97.60% paling optimal dalam pengujian data</p>
2	<p>Firman Dani, Asih Asmarani, Nova Selvia / 2022 [20]</p>	<p>Deteksi Serangan Web Attack SQL Injection Menggunakan Algoritma C4.5 dan Naïve Bayes</p>	<p>Algoritma C4.5 dan Naïve Bayes</p>	<p>Performa algoritma <i>Naïve Bayes</i> lebih baik dalam mengidentifikasi jenis trafik</p>

				serangan <i>sql injection</i> pada trafik jaringan, sedangkan C4.5 unggul dalam pengklasifikasian trafik normal.
3	Sari Sandra, Deris Stiawan, Ahmad Heryanto / 2016 [21]	Visualisasi Serangan Brute Force Menggunakan Metode K-Means dan Naïve Bayes	K-Means dan Naïve Bayes	Metode K-Means dan metode Naïve Bayes dapat diimplementasikan pada <i>dataset</i> dalam mengkategorikan sejumlah paket data <i>attack</i> atau paket data normal berdasarkan <i>attack</i> dan <i>normal pattern</i> . Hasil akhir dari implementasi

				<p>kedua metode dapat memberikan visualisasi dalam bidang <i>two dimensional</i> (2D), berupa visualisasi <i>scatter plot</i> atau <i>parallel coordinate</i>, dengan <i>accuracy</i> pengkategorian yang baik.</p> <p>Metode K-Means dan metode Naïve Bayes yang diimplementasikan pada <i>ISCX dataset</i> mendapatkan hasil <i>accuracy</i> hingga 95,46% dan 99,68%,</p>
--	--	--	--	--

				sedangkan pada DARPA <i>dataset</i> didapatkan nilai <i>accuracy</i> 73,60% dan 98,79%.
4	Kurniabudi, Abdul Harris, Albertus Edward Mintaria / 2021 [22]	Komparasi <i>Information Gain</i> , <i>Gain Ratio</i> , <i>CFs-Bestfirst</i> dan <i>CFs-PSO Search</i> Terhadap Performa Deteksi Anomali	Naive Bayes, k-NN, dan J48	Pengujian memperlihatkan bahwa teknik seleksi fitur <i>IG</i> , <i>GR</i> , <i>CB</i> dan <i>CF-PSO</i> menghasilkan nilai bobot yang berbeda, dimana hal ini juga berdampak pada jenis dan jumlah fitur yang dihasilkan. Pada teknik seleksi fitur <i>IG</i> dan <i>GR</i>

				<p>diperlukan intervensi pakar untuk menentukan batas minimum bobot fitur, dimana batas minimum ini mempengaruhi jumlah dan jenis fitur yang akan digunakan pada algoritma klasifikasi untuk mendeteksi serangan. Hasil pengujian menunjukkan teknik seleksi fitur yang penulis gunakan seperti <i>IG</i>, <i>GR</i>, <i>CB</i> dan <i>CF-PSO</i> mampu</p>
--	--	--	--	---

				<p>meningkatkan algoritma klasifikasi <i>NB</i>, <i>k-NN</i> dan <i>J48</i> untuk mengklasifikasikan trafik normal dan serangan tertentu, meskipun belum maksimal.</p>
--	--	--	--	--

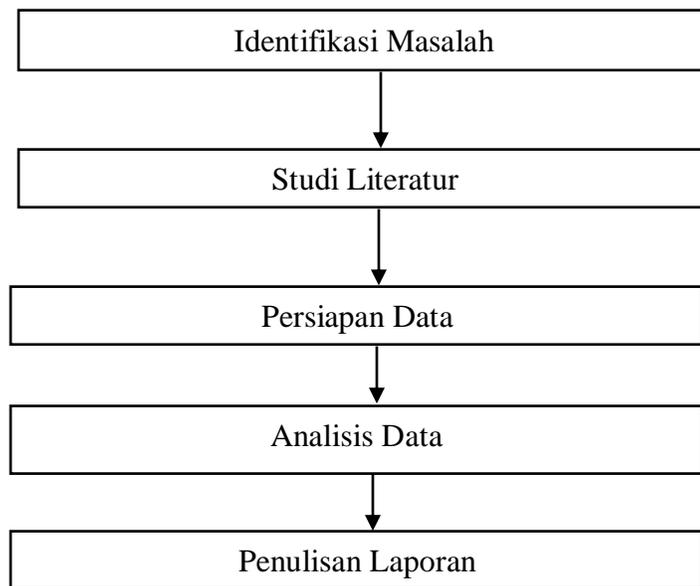
Dari beberapa penelitian diatas, terdapat berapa kesamaan dengan penelitian yang akan dilakukan yaitu metode klasifikasi digunakan untuk pengidentifikasian suatu data, penelitian yang dibuat memiliki masalah yang sama yaitu tentang deteksi serangan, penelitiannya sama-sama mengangkat topik *data mining*. Akan tetapi terdapat juga perbedaan antara penelitian sejenis dengan penelitian yang akan dibuat diantaranya yaitu objek atau *dataset* yang digunakan berbeda, jumlah *dataset* yang dipakai, fitur yang digunakan, jumlah anomali yang dideteksi serta perbandingan komparasi algoritma yang akan digunakan dalam penelitian. Sehingga peneliti menyimpulkan bahwa perlu dilakukan penelitian lebih lanjut untuk menguji performa algoritma yang baik apabila digunakan dalam mendeteksi beberapa serangan.

BAB III

METODOLOGI PENELITIAN

3.1 KERANGKA KERJA PENELITIAN

Untuk membantu penyusunan dan memperjelas tahapan yang akan dilakukan pada penelitian, maka perlu dibuat kerangka kerja penelitian guna memperjelas tahapan-tahapan yang akan dilakukan dalam menyelesaikan permasalahan yang ada. Adapun kerangka kerja penelitian dapat dilihat pada gambar 3.1.



Gambar 3.1 Kerangka Kerja Penelitian

Berdasarkan kerangka kerja penelitian diatas maka dapat diuraikan pembahasan masing-masing tahapan sebagai berikut :

3.1.1 Identifikasi Masalah

Identifikasi permasalahan dalam penelitian ini adalah mendeteksi data serangan atau anomali trafik pada trafik jaringan, serta mencari metode yang memiliki performa yang baik dalam mengidentifikasi anomali trafik/serangan. Pada penelitian ini peneliti menggunakan algoritma *Naïve Bayes* dan *Random Forest* dalam pengujian pendeteksian serangan.

3.1.2 Studi Literatur

Pada tahapan ini dilakukan kajian pustaka yaitu pencarian landasan-landasan teori yang relevan dengan permasalahan yang diteliti, yang dapat diperoleh dari buku, jurnal, internet seperti *google scholar*, *ebook*, dll.

3.1.3 Persiapan Data

Adapun tahapan *preprocessing* data yang akan dilakukan mencakup beberapa proses meliputi :

a. Pemilihan *Dataset*

Pada penelitian ini *dataset* yang digunakan merupakan data publik yaitu *dataset* CICIDS2017 dari ISCX Consortium yang dapat diakses pada laman www.unb.ca. *Dataset* yang dipakai merupakan salah satu dari *dataset* CICIDS2017 yaitu Thursday-WorkingHours-Morning-WebAttack.pcap_ISCX yang didalamnya terdiri dari trafik normal/*benign*, dan trafik anomali seperti trafik *WebAttack Brute Force*, *WebAttack Sql Injection* dan *WebAttack XSS*.

b. Pembersihan Data (*Data Cleaning*)

Pada tahapan ini dilakukan pembersihan data dan memperkecil data yang tidak konsisten atau data yang tidak relevan seperti menghilangkan data atau atribut yang berpengaruh pada proses klasifikasi, biasa mencakup membuang duplikasi data, memeriksa data yang tidak konsisten serta memperbaiki kesalahan data.

c. *Data Split*

Dimana pada tahapan ini mempersiapkan data yang akan digunakan untuk dianalisa, tahap ini data dibagi menjadi dua bagian berupa data *training* dan data *testing*. Perbandingan yang dipakai pada penelitian ini yaitu 70 : 30 yang berarti dari keseluruhan dataset yang dipakai 70% akan menjadi data *training* dan 30% sebagai data *testing*.

d. *Data Reduction* (Pemilihan Fitur)

Data reduction dilakukan untuk mengurangi dimensi atau fitur yang tidak relevan. Pada penelitian ini menggunakan seleksi fitur *information gain*.

e. *Data Transformasi*

Dimana setelah dilakukan tahap seleksi data, selanjutnya dilakukan pengubahan data ke format yang sesuai untuk kemudian diproses dalam tahapan *data mining*. Pada penelitian ini data diubah menjadi format *file .Arff* (*Attribute Relation File Format*) agar dapat digunakan pada *tools Weka*.

3.1.4 Analisis Data

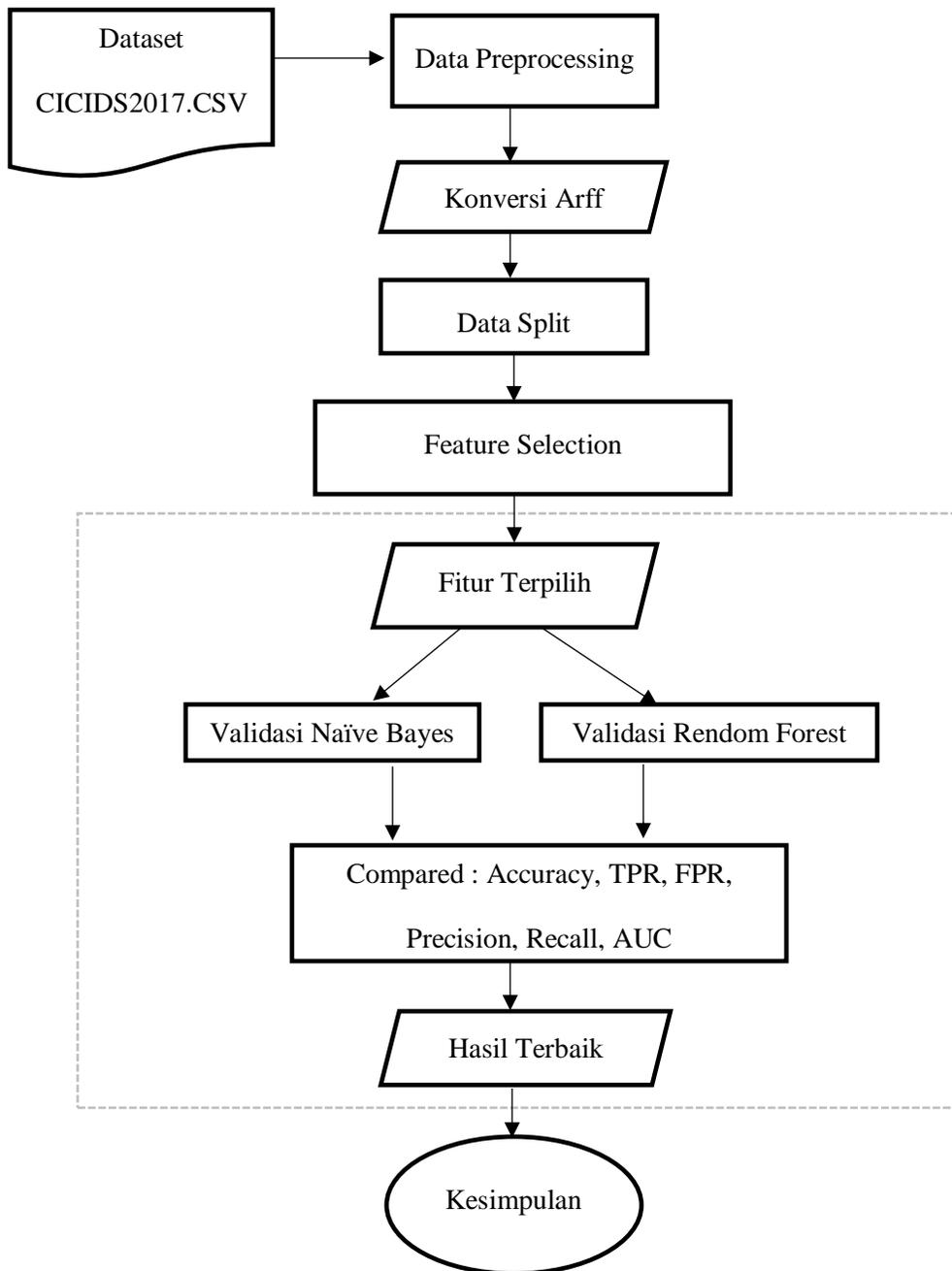
Pada tahapan ini data dilakukan tahap validasi atau pengklasifikasian menggunakan algoritma *Naïve Bayes* dan *Random Forest*, kemudian dilakukan evaluasi terhadap keefektifan dan kualitas algoritma yang digunakan sehingga menghasilkan hasil akurasi untuk masing-masing algoritma klasifikasi yang dipakai.

3.1.5 Penulisan Laporan

Setelah semua tahapan penelitian dilakukan, selanjutnya pada tahapan ini dari hasil eksperimen yang dilakukan maka dibuat laporan penelitian terhadap permasalahan yang diangkat serta hasil kesimpulan dari penelitian yang telah dilakukan sebagai dokumentasi penelitian.

3.2 RANCANGAN EKSPERIMEN

Adapun rancangan alur eksperimen yang akan dilakukan yaitu sebagai berikut :



3.3 ALAT BANTU PENELITIAN

Beberapa alat bantu (*tools*) yang digunakan dalam penelitian ini sebagai berikut:

1. Perangkat Keras (*Hardware*)

Berikut perangkat keras yang digunakan diantaranya yaitu :

- a. Processor : AMD A10-7300 Radeon R6, 10 Compute Core 4C+6G
1.90GHz
- b. Memory : 4 GB
- c. Hardisk : 1 TB
- d. Printer EPSON L220

2. Perangkat Lunak (*Software*)

Berikut perangkat lunak yang digunakan diantaranya yaitu :

- a. Sistem Operasi Windows 8.1 Pro
- b. Microsoft Word 2013
- c. Microsoft Excel 2013
- d. Mendeley
- e. Weka
- f. Notepad
- g. Google Chrome
- h. Adobe Reader

BAB IV
JADWAL PENELITIAN

Estimasi waktu dilakukannya penelitian ini yaitu pada bulan Oktober 2022 sampai Januari 2022.

Kegiatan	Oktober				November				Desember				Januari			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Identifikasi Masalah																
Studi Literatur																
Persiapan Data																
Perhitungan dan Analisis																
Evaluasi Data																
Penulisan Hasil Laporan																

DAFTAR PUSTAKA

- [1] M. Anif, S. Hws, and M. D. Huri, “Penerapan Intrusion Detection System (IDS) dengan metode Deteksi Port Scanning pada Jaringan Komputer di Politeknik Negeri Semarang,” vol. 13, pp. 25–30, 2015.
- [2] M. K. Harto and A. Basuki, “Deteksi Serangan DDoS Pada Jaringan Berbasis SDN Dengan Klasifikasi Random Forest,” vol. 5, no. 4, pp. 1329–1333, 2021.
- [3] M. Fadhlurrohman, A. Muliawati, and B. Hananto, “Analisis Kinerja Intrusion Detection System pada Deteksi Anomali dengan Metode Decision Tree Terhadap Serangan Siber Analysis of Intrusion Detection System Performance on Anomaly Detection with Decision Tree Method Against Cyber Attacks,” vol. 8, no. Pratomo 2016, pp. 90–94.
- [4] S. K. AM Siregar, S Kom, MKDANA Puspabhuana, *DATA MINING : Pengolahan Data Menjadi Informasi dengan RapidMiner*. 2017.
- [5] N. A. Amalia Rizqi Utami, Yudha Purwanto, “PENGELOMPOKAN TRAFIK BERDASARKAN WAKTU DENGAN ALGORITMA CLUSTREAM UNTUK DETEKSI ANOMALI PADA ALIRAN TRAFIK TIME BASED TRAFFIC CLUSTERING USING CLUSTREAM ALGORITHM FOR ANOMALY DETECTION ON,” vol. 4, no. 1, pp. 848–854, 2017.
- [6] M. A. Shauma, Y. Purwanto, and A. Novianty, “Deteksi Anomali Trafik

Menggunakan Algoritma Birch Dan Dbscan Pada Streaming Traffic,” *eProceedings Eng.*, vol. 3, no. 3, pp. 5004–5012, 2016, [Online]. Available: <https://librarye proceeding.telkomuniversity.ac.id/index.php/engineering/article/view/3132>.

- [7] S. R. Niko Suwaryo¹, Ismasari Nawangsih², “Deteksi Serangan pada Intrusion Detection System (IDS) untuk Klasifikasi Serangan dengan Algoritma Naïve Bayes, C.45 dan K-NN dalam Meminimalisasi Resiko Terhadap Pengguna,” *Angew. Chemie Int. Ed.* 6(11), 951–952., pp. 2013–2015, 2021.
- [8] I. Sumarno and M. M. S. Bisosro, “Solusi Network Security Dari Ancaman Sql Injection Dan Denial of Service (Dos),” *Teknolojia*, vol. 5, pp. 19–30, 2003, [Online]. Available: <http://journal.umsida.ac.id/files/Sumarno Rev.pdf>.
- [9] G. Meena, “2017 International Conference on Computer, Communications and Electronics, COMPTELIX 2017,” *2017 Int. Conf. Comput. Commun. Electron. COMPTELIX 2017*, pp. 553–558, 2017.
- [10] Jupriyadi, “Implementasi Seleksi Fitur Menggunakan Algoritma FVBRM Untuk Klasifikasi Serangan Pada Intrusion Detection System (Ids),” *Semin. Nas. Teknol. Inf.*, vol. 17, no. January 2018, pp. 1–6, 2018, [Online]. Available: <https://jurnal.umj.ac.id/index.php/semnastek/article/view/3452/2601>.
- [11] I. Sharafaldin, “Intrusion Detection Evaluation Dataset (CIC-IDS2017),”

www.unb.ca, 2018. <https://www.unb.ca/cic/datasets/ids-2017.html> (accessed May 20, 2022).

- [12] Kurniabudi, D. Stiawan, Darmawijoyo, M. Y. Bin Bin Idris, A. M. Bamhdi, and R. Budiarto, "CICIDS-2017 Dataset Feature Analysis with Information Gain for Anomaly Detection," *IEEE Access*, vol. 8, pp. 132911–132921, 2020, doi: 10.1109/ACCESS.2020.3009843.
- [13] M. S. Indah Werdiningsih, S.Si., M.Kom., Barry Nuqoba, S.Si., M.Kom., Muhammadun, S.Si., *Data Mining Menggunakan Android, Weka, dan SPSS*. Jawa Timur: Airlangga University Press, 2020.
- [14] Y. D. Atma and A. Setyanto, "Perbandingan algoritma c4.5 dan k-nn dalam identifikasi mahasiswa berpotensi drop out," *Metik J. ISSN 2580-1503*, vol. 2, no. 2, pp. 31–37, 2018.
- [15] R. K. Niswatin, "Sistem Pendukung Keputusan Penempatan Jurusan Mahasiswa Baru Menggunakan Metode K-nearest Neighbor," *Cogito Smart J.*, vol. 1, no. 1, pp. 55–67, 2015, doi: 10.31154/cogito.v1i1.6.55-67.
- [16] F. Gorunescu, *Data Mining: Concepts, Models, and Techniques*, Springer, Verlag Berlin Heidelberg, 2011.
- [17] E. Prasetyo, *Data Mining - Konsep dan Aplikasi Menggunakan MATLAB*. Penerbit ANDI, 2012.

- [18] M. Faid, “Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi,” vol. 8, 2019, doi: 10.34148/teknika.v8i1.95.
- [19] Ainurrohmah, “Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka,” *Prisma*, vol. 4, pp. 493–499, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>.
- [20] N. S. Firman Dani, Asih Asmarani, “Deteksi Serangan WebAttack SQL Injection Menggunakan Algoritma C4.5 dan Naive Bayes,” 2022.
- [21] S. Sandra, D. Stiawan, and A. Heryanto, “Visualisasi Serangan Brute Force Menggunakan Metode K-Means dan Naïve Bayes,” *Proceeding - Annu. Res. Semin. Proceeding*, vol. 2, no. 1, pp. 315–320, 2016.
- [22] K. Kurniabudi, A. Harris, and A. E. Mintaria, “Komparasi Information Gain, Gain Ratio, CFs-Bestfirst dan CFs-PSO Search Terhadap Performa Deteksi Anomali,” *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 332, 2021, doi: 10.30865/mib.v5i1.2258.

LAMPIRAN DATASET

```
Thursday-WorkingHours-Morning-WebAttacks.peap_ISCX.aff
1 @relation Thursday-WorkingHours-Morning-WebAttacks.peap_ISCX
2
3 @attribute ' Destination Port' numeric
4 @attribute ' Flow Duration' numeric
5 @attribute ' Total Fwd Packets' numeric
6 @attribute ' Total Backward Packets' numeric
7 @attribute 'Total Length of Fwd Packets' numeric
8 @attribute ' Total Length of Bwd Packets' numeric
9 @attribute ' Fwd Packet Length Max' numeric
10 @attribute ' Fwd Packet Length Min' numeric
11 @attribute ' Fwd Packet Length Mean' numeric
12 @attribute ' Fwd Packet Length Std' numeric
13 @attribute 'Bwd Packet Length Max' numeric
14 @attribute ' Bwd Packet Length Min' numeric
15 @attribute ' Bwd Packet Length Mean' numeric
16 @attribute ' Bwd Packet Length Std' numeric
17 @attribute 'Flow Bytes/s' numeric
18 @attribute ' Flow Packets/s' numeric
19 @attribute ' Flow IAT Mean' numeric
20 @attribute ' Flow IAT Std' numeric
21 @attribute ' Flow IAT Max' numeric
22 @attribute ' Flow IAT Min' numeric
23 @attribute 'Fwd IAT Total' numeric
24 @attribute ' Fwd IAT Mean' numeric
25 @attribute ' Fwd IAT Std' numeric
26 @attribute ' Fwd IAT Max' numeric
27 @attribute ' Fwd IAT Min' numeric
28 @attribute 'Bwd IAT Total' numeric
29 @attribute ' Bwd IAT Mean' numeric
30 @attribute ' Bwd IAT Std' numeric
31 @attribute ' Bwd IAT Max' numeric
32 @attribute ' Bwd IAT Min' numeric
33 @attribute 'Fwd PSH Flags' numeric
34 @attribute ' Bwd PSH Flags' numeric
35 @attribute ' Fwd URG Flags' numeric
36 @attribute ' Bwd URG Flags' numeric
37 @attribute ' FwdHeaderLength' numeric
```

Normal text file
length: 50,662,445 lines: 170,450 Ln: 1 Col: 1 Pos: 1
UTF-8 Unix (LF) INS

38	@attribute	' Bwd Header Length'	numeric
39	@attribute	' Fwd Packets/s'	numeric
40	@attribute	' Bwd Packets/s'	numeric
41	@attribute	' Min Packet Length'	numeric
42	@attribute	' Max Packet Length'	numeric
43	@attribute	' Packet Length Mean'	numeric
44	@attribute	' Packet Length Std'	numeric
45	@attribute	' Packet Length Variance'	numeric
46	@attribute	' FIN Flag Count'	numeric
47	@attribute	' SYN Flag Count'	numeric
48	@attribute	' RST Flag Count'	numeric
49	@attribute	' PSH Flag Count'	numeric
50	@attribute	' ACK Flag Count'	numeric
51	@attribute	' URG Flag Count'	numeric
52	@attribute	' CWE Flag Count'	numeric
53	@attribute	' ECE Flag Count'	numeric
54	@attribute	' Down/Up Ratio'	numeric
55	@attribute	' Average Packet Size'	numeric
56	@attribute	' Avg Fwd Segment Size'	numeric
57	@attribute	' Avg Bwd Segment Size'	numeric
58	@attribute	' Fwd Header Length'	numeric
59	@attribute	' Fwd Avg Bytes/Bulk'	numeric
60	@attribute	' Fwd Avg Packets/Bulk'	numeric
61	@attribute	' Fwd Avg Bulk Rate'	numeric
62	@attribute	' Bwd Avg Bytes/Bulk'	numeric
63	@attribute	' Bwd Avg Packets/Bulk'	numeric
64	@attribute	' Bwd Avg Bulk Rate'	numeric
65	@attribute	' Subflow Fwd Packets'	numeric
66	@attribute	' Subflow Fwd Bytes'	numeric
67	@attribute	' Subflow Bwd Packets'	numeric
68	@attribute	' Subflow Bwd Bytes'	numeric
69	@attribute	' Init Win bytes forward'	numeric
70	@attribute	' Init Win bytes backward'	numeric
71	@attribute	' act_data_pkt_fwd'	numeric
72	@attribute	' min_seg_size_forward'	numeric
73	@attribute	' Active Mean'	numeric
74	@attribute	' Active Std'	numeric

