BAB II
TINJAUAN PUSTAKA

2.1. Studi Literatur

Di dalam landasan teori akan dibahas dahulu ringkasan studi literatur yang dilakukan untuk mengetahui sejauh mana penelitian tentang algoritma *Machine Learning* serta serangan-serangan DDoS dilakukan

Tabel 2.1 Studi Literatur

Nama Peneliti	**Judul Penelitian**	**Metode Deteksi**	**Tahun Publikasi**
- Solomon Mwanjele - Masinde Muthoni - Peter Ochieg	Comparison of Nearest Neighbor (IBk), Regression by Discretization and Isotonic Regression Classification Algorithms for Precipitation Classes Prediction	Nearest Neighbor (IBk), Regression By Discretization, Isotonic Regression	2014
- Dr. Vaishali S. Parsania - Dr. N. N. Jani - Navneet H Bhalodiya	Applying Naïve Bayes, BayesNet, PART, JRip and OneR Algorithms on Hypothyroid Database for Comparative Analysis	Naïve Bayes, Bayesnet, PART, Jrip, OneR	2014
- Sulidar Fitri	Perbandingan Kinerja Algoritma Klasifikasi Naïve Bayesian, Lazy-IBk, Zero-R, Dan Decision Tree- J48	Naïve Bayesian, Lazy-IBk, Zero-R, Decision Tree- J48	2014

Lanjutan Tabel 2.1 Studi Literatur

Nama Peneliti	**Judul Penelitian**	**Metode Deteksi**	**Tahun Publikasi**
- K.R.W.V. Bandara - T.S. Abeysinghe - A.J. M. Hijaz - D.G.T. Darshana - H. Aneez - S.J. Kaluarachchi - K.V.D.L. Sulochana - Mr. Dhishan Dhammearatchi	Preventing DDoS attack using Data mining Algorithms	C.4.5 (Decision Tree)	2015
- Ashish Kumar Dogra, - Tanuj Wala	A Comparative Study of Selected Classification Algorithms of Data Minin	Zero-R, Part, Prism, OneR, J48	2015
- Irina Pak - Phoey Lee Teh	*Machine Learning* Classifiers: Evaluation of the Performance in Online Reviews	Bayes, Functions, Lazy, Meta, Rules, Trees.	2016
- Anand Kishor Pandey - Dharmveer Singh Rajpoot	A comparative study of classification techniques by utilizing WEKA	Decision Stump, Random Forest, J48, NaiveBayes, NaiveBayes Simple, BayesNet	2016

Lanjutan Tabel 2.1 Studi Literatur

Nama Peneliti	**Judul Penelitian**	**Metode Deteksi**	**Tahun Publikasi**
- Anton Yudhana - Imam Riadi - Faizin Ridho	DDoS Classification Using Neural Network and Naïve Bayes Methods for Network Forensics	Neural Network, Naïve Bayes	2018
- Hilman Nihri - Eko Sakti Pramukantoro - Primantara Hari Trisnawan	Pengembangan IDS Berbasis J48 Untuk Mendeteksi Serangan DoS Pada Perangkat Middleware IoT	J48 (Decision Tree)	2018
- Santosh Kumar Pydipalli, - Srikanth Kasthuri - Jinu S	DDoS Detection System Using C4.5 Decision Tree Algorithm	J48 (Decision Tree)	2018
- Hariharan. M, Abhishek H. K - B. G. Prasad	Ddos Attack Detection Using C5.0 *Machine Learning* Algorithm	C5.0 (Decision Tree)	2019

2.2. *Machine Learning*

Pembelajaran mesin (*Machine Learning*) adalah pengaplikasian Kecerdasan Buatan (*Artificial Intelligence*) yang menyediakan program komputer yang memilik kemampuan untuk belajar dari pengimputan data [4]. Hal ini memungkinkan kita untuk menggunakan data historis sebagai input untuk prediksi data masa depan. Dengan demikian, akurasi output semata-mata didasarkan pada kualitas data historis.

Saat ini, teknik pembelajaran mesin digunakan di berbagai bidang untuk memecahkan masalah yang berbeda. Misalnya, mereka digunakan dalam penyaringan spam e-mail, pengenalan pola dan gambar, penyaringan mesin pencari, aplikasi perawatan kesehatan, dll.

2.2.1. Tipe Algoritma *Machine Learning*

Algoritma pembelajaran mesin dapat secara luas diklasifikasikan sebagai Algoritma pembelajaran terawasi (*Supervised Learning*) dan Algoritma pembelajaran tanpa pengawasan (*Unsupervised learning*) [4].

a. Algoritma Pembelajaran Terawasi

   Algoritma pembelajaran yang diawasi (*Supervised Learning*) terutama digunakan untuk menyelesaikan masalah klasifikasi dan regresi karena membuat proses pendeteksian atau pengambilan keputusan menjadi lebih mudah. Ini menggunakan data yang dipelajari di masa lalu untuk memprediksi peristiwa di masa depan. Data input yang digunakan untuk melatih algoritma pembelajaran adalah yang berlabel [4].

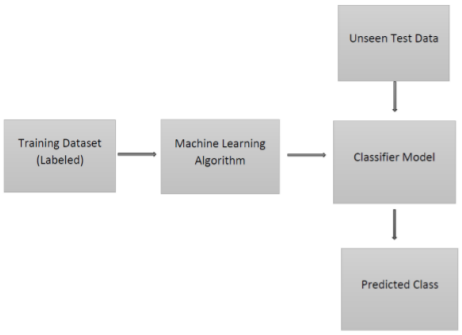
b. Algoritma Pembelajaran Tanpa Pengawasan

   Algoritma pembelajaran tanpa pengawasan (*Unsupervised learning*) menggunakan data input yang tidak berlabel untuk melatih sistem. Artinya, data input tidak ditandai dengan label. Ia menemukan struktur tersembunyi dari input yang tidak berlabel, dan mengelompokkannya sebagai kelompok yang menunjukkan kesamaan. Kinerja awal dari jenis algoritma pembelajaran ini buruk, tetapi sistem dapat menyesuaikan dirinya untuk meningkatkan kinerja [4].

2.2.2. *Supervised Machine Learning*

Algoritma pembelajaran yang diawasi (*Supervised Learning*) adalah teknik yang paling umum digunakan dalam pembelajaran mesin [4]. Masalah penelitian ini mengimplementasikan algoritma pembelajaran mesin yang dilindungi untuk mengklasifikasikan lalu lintas jaringan sebagai lalu lintas yang sah dan lalu lintas beberapa. Di sink *classifier* mendapatkan input yang merupakan seperangkat nilai

fitur yang juga disebut *Vector Input* dan menampilkan nilai prediksi yang disebut kelas. Gambar 2.1 menunjukkan klasifikasi dari *Supervised Learning*.



Gambar 2.1 Klasifikasi *Supervised Learning*

Di sini data pelatihan diberikan sebagai *input* ke algoritma pembelajaran yang menghasilkan model *classifier*. Kinerja dari *classifier* dapat dievaluasi menggunakan data yang tidak terlihat.

2.3. Klasifikasi

Decision Tree, K-NN(IBk), Bayesian Network dan *OneR* digunakan sebagai algoritma klasifikasi untuk memprediksi hasil analisa dari klasifikasi serangan DDoS di jaringan.

2.3.1. *Bayesian Network*

Jaringan *Bayesian* adalah struktur yang menunjukkan dependensi kondisional antara variabel domain dan juga dapat digunakan untuk menggambarkan secara grafis hubungan yang mendasari probabilistik antara variabel domain. Jaringan Bayesian terdiri dari grafik asiklik terarah dan tabel probabilitas. *Node* jaringan mewakili variabel domain dan busur antara dua node menunjukkan adanya hubungan yang mendasari atau ketergantungan antara dua node ini [5].

Metode *Bayesian Network* merupakan metode yang baik di dalam *Machine Learning* berdasarkan data *training*, dengan menggunakan probabilitas bersyarat sebagai dasarnya. *Bayesian Network* terdiri dari dua bagian utama, yaitu:

a. Struktur *graf Bayesian Network* disebut dengan *Directed Acyclic Graph (DAG*). DAG terdiri dari *node* dan *edge*. *Node* merepresentasikan variabel acak dan *edge* merepresentasikan adanya hubungan ketergantungan langsung

dan dapat juga diinterprestasikan sebagai pengaruh (sebab-akibat) antara variabel yang dihubungkannya. Tidak adanya *edge* menandakan adanya hubungan kebebasan kondisional di antara variabel.

b. Himpunan parameter, mendefinisikan distribusi probabilitas kondisional untuk setiap variabel. Pada *Bayesian Network*, *node* berkorespondensi dengan variabel acak. Tiap *node* diasosiasikan dengan sekumpulan peluang bersyarat, p(Xi|Ai) sehingga Xi adalah variabel yang diasosiasikan dengan *node* dan Ai adalah set dari *parent* dalam *graph*.

Dalam membangun *Bayesian Network*, struktur dibangun dengan pendekatan statistik yang dikenal dengan *Theorema Bayes* yaitu *Conditional Probability* (peluang bersyarat). *Conditional Probability* yaitu perhitungan peluang suatu kejadian B bila diketahui kejadian A telah terjadi, dinotasikan dengan P(A). *Teorema* ini digunakan untuk menghitung peluang suatu set data untuk masuk ke dalam suatu kelas tertentu berdasarkan inferensi data yang sudah ada. Adapun rumus dasar dari *Theorema Bayes*, yaitu:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \quad (1)$$

Atau menggunakan rumus dibawah ini:

$$P(A|B) = \frac{P(A)\,P(B|A)}{P(B|A)\,P(A) + P(B|A)\,P(A)} \quad (2)$$

2.3.2. *OneR*

OneR, adalah algoritma klasifikasi sederhana yang menghasilkan pohon keputusan satu tingkat. *OneR* mampu menyimpulkan aturan klasifikasi yang sederhana, namun tepat, dari satu set instance. *OneR* juga mampu menangani nilai-nilai yang hilang dan atribut numerik yang menunjukkan fleksibilitas terlepas dari kesederhanaan. Algoritma *OneR* membuat satu aturan untuk setiap atribut dalam data pelatihan, lalu memilih aturan dengan tingkat kesalahan minimum sebagai "satu aturan" nya [5].

2.3.3. *Decision Tree* Menggunakan J48

Decision tree Ini adalah termasuk kedalam algoritma pembelajaran yang dilindungi. Aturan *Decision Tree* (pohon keputusan) mudah dipahami pengguna

dan menggunakan sistem pengetahuan seperti alat Weka. C4.5 disebut sebagai (J48 dalam perangkat lunak Weka). Motif utama menggunakan aturan pohon keputusan mereka adalah untuk membuat model pelatihan dan yang diprediksi nilai kelasnya. Di sini informasi mendapatkan rasio sebagai jumlah untuk memilih fitur pemisahan. Pohon keputusan diklasifikasikan ke dalam struktur pohon, pohon berisi simpul keputusan dan simpul daun [6].

Algoritma J48 sendiri adalah sebuah algoritma turunan dari C4.5. Algoritma ini menghasilkan pohon *biner* dimana dalam proses klasifikasi pohon akan dibangun dan setiap *tupel* dari pohon tersebut akan diterapkan pada basis data dan hasil klasifikasi dari tupel tersebut [7]. Algoritma J48 akan mengabaikkan nilai yang tidak lengkap dalam proses pembuatan pohon. Dasar dari algoritma ini adalah untuk membagi data ke dalam beberapa bagian berdasarkan nilai atribut dari item yang ada pada training *dataset*. Algoritma J48 dapat melakukan klasifikasi baik melalui *decision tree* ataupun *rules* yang diperoleh dari pohon tersebut. Adapun langkah-langkah dalam algoritma J48 adalah:

a. Menetapkan training *dataset*.

b. Penentuan akar dari pohon keputusan.

c. Penghitungan nilai *Gain* menggunakan persamaan:

$$Entropy(S)=\sum_{i=1}^{n} -pi * log_2\ pi \quad (3)$$

Mulai kembali langkah ke-2 sampai semua *tupel* terbagi dengan menggunakan persamaan:

$$Gain\ (S, A) = S - \sum_{i=1}^{n} -\frac{|\ s_i\ |}{|\ s\ |} * S_1 \quad (4)$$

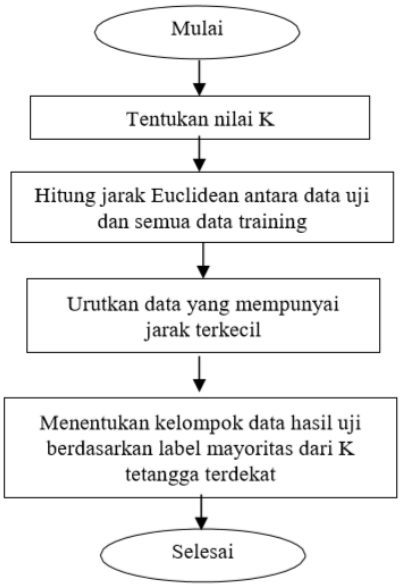
d. Proses pembagian akan berhenti ketika semua *tupel* dalam titik N telah memperoleh kelas yang sama dana tau sudah tidak ada atribut dalam *tupel* yang dibagi lagi atau tidak ada *tupel* dalam cabang yang kosong.

Untuk mengoptimasi parameter klasifikasi digunakan 10-fold validasi silang (cross validation). Kemudian testing *dataset* digunakan untuk mengukur dan menguji validitas model prediksi yang dikembangkan. Hasil akhir dari tahapan-tahapan di atas adalah berupa model prediksi akhir yang dapat memberikan pengetahuan baru.

2.3.4. *K-NN (K-Nearest Neighbors)*

K-Nearest Neighbor (KNN) adalah metode melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Metode ini bertujuan untuk mengklasifikasikan objek baru berdasarkan atribut dan training *sample*. Diberikan suatu titik *query*, selanjutnya akan ditemukan sejumlah K objek atau titik training yang paling dekat dengan titik *query*. Nilai prediksi dari *query* akan ditentukan berdasarkan klasifikasi tetanggaan [8].

Sebelum melakukan perhitungan dengan metode *K-Nearest Neighbors*, terlebih dahulu harus menentukan data latih dan data uji. Kemudian akan dilakukan proses perhitungan untuk mencari jarak menggunakan *Euclidean*. Setelah itu, akan dilakukan tahapan perhitungan dengan metode KNN seperti pada Gambar 2.2.



Gambar 2.2 Proses Metode *K-Nearest Neighbors*

Algoritma *K-Nearest Neighbor (KNN)* juga merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Teknik ini sangat sederhana dan mudah diimplementasikan. Mirip dengan teknik *clustering*, yaitu mengelompokkan suatu data baru berdasarkan jarak data baru itu ke beberapa data/tetangga terdekat. Pertama sebelum mencari jarak data ke tetangga adalah menentukan nilai K tetangga (*neighbor*). Lalu, untuk mendefinisikan jarak antara dua titik yaitu titik

pada data training dan titik pada data testing, maka digunakan rumus *Euclidean* dengan persamaan sebagai berikut:

$$d(a,b) = \sum_{i=0}^{n} (Xi - Yi)^2 \quad (5)$$

Dengan keterangan:

d (a,b) : jarak *Euclidian*	y : data 2	n : jumlah fitur
x : data 1	i : fitur ke –	

2.4. *Distributed Denial of Service (DDoS)*

DDoS merupakan kependekan dari *Distributed Denial of Service* atau dalam bahasa Indonesia dapat diartikan sebagai Penolakan Layanan secara Terdistribusi. DDoS adalah jenis serangan yang dilakukan dengan cara membanjiri lalu lintas jaringan internet pada server, sistem, atau jaringan. Umumnya serangan ini dilakukan menggunakan beberapa komputer *host* penyerang sampai dengan komputer target tidak bisa diakses [2].

Ada sejumlah studi survei yang telah mengusulkan taksonomi sehubungan dengan serangan DDoS [3]. Meskipun semua telah melakukan pekerjaan yang terpuji dalam mengusulkan *taksonomi* baru, cakupan serangan sejauh ini terbatas. Ada kebutuhan untuk mengidentifikasi serangan baru dan menghasilkan *taksonomi* baru. Oleh karena itu, kami telah menganalisis serangan baru yang dapat dilakukan menggunakan protokol berbasis TCP / UDP di lapisan aplikasi dan mengusulkan taksonomi baru. Sisa dari sub-bagian ini telah dijelaskan *taksonomi* terperinci dari serangan DDoS dan diilustrasikan dalam Gambar 2.3, dalam hal serangan berbasis refleksi dan eksploitasi.

4.1.1. Serangan DDoS Berbasis Refleksi

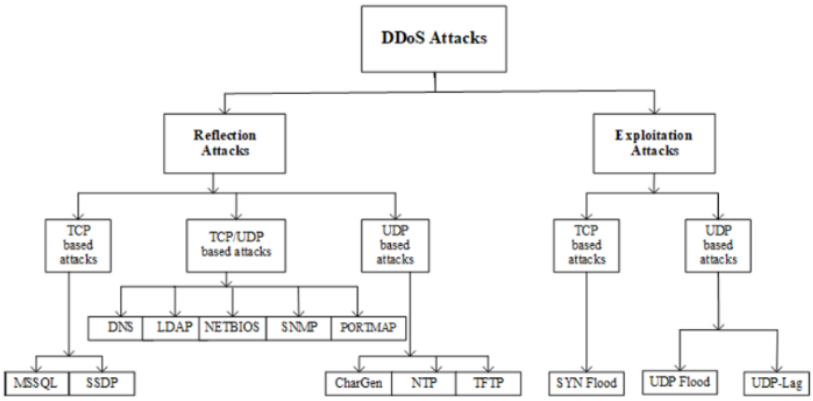
Serangan DDoS berbasis refleksi: Apakah jenis-jenis serangan di mana identitas penyerang tetap tersembunyi dengan menggunakan komponen pihak ketiga yang sah. Paket-paket dikirim ke server *reflektor* oleh penyerang dengan alamat IP sumber diatur ke alamat IP target korban untuk membanjiri korban dengan paket respon[3]. Serangan-serangan ini dapat dilakukan melalui protokol *layer*

application menggunakan protokol *layer transport*, yaitu protokol kontrol transmisi (TCP), protokol datagram pengguna (UDP) atau melalui kombinasi keduanya. Seperti yang ditunjukkan Gambar diatas, dalam kategori ini, serangan berbasis TCP termasuk MSSQL, SSDP sedangkan serangan berbasis UDP termasuk CharGen, NTP dan TFTP. Ada beberapa serangan yang dapat dilakukan dengan menggunakan TCP atau UDP seperti DNS, LDAP, NETBIOS, dan SNMP.

2.4.1. Serangan DDoS berbasis Eksploitasi

Serangan DDoS berbasis eksploitasi: Apakah **serangan** semacam itu di mana identitas penyerang tetap disembunyikan dengan menggunakan komponen pihak ketiga yang sah. Paket-paket dikirim ke *server reflektor* oleh penyerang dengan alamat IP sumber diatur ke alamat IP target korban untuk membanjiri korban dengan paket tanggapan [3]. Serangan-serangan ini juga dapat dilakukan melalui protokol layer aplikasi menggunakan protokol *transport layer* misalnya TCP dan UDP. Serangan eksploitasi berbasis TCP termasuk banjir SYN dan serangan berbasis UDP termasuk banjir UDP dan UDP-Lag. Serangan banjir UDP dimulai pada host jarak jauh dengan mengirimkan sejumlah besar paket UDP.

Paket-paket UDP ini dikirim ke port acak pada mesin target pada tingkat yang sangat tinggi. Akibatnya, *bandwidth* jaringan yang tersedia menjadi habis, sistem macet dan kinerja menurun. Di sisi *lane*, banjir SYN juga mengkonsumsi sumber daya *server* dengan mengeksploitasi jabat tangan tiga arah TCP. Serangan ini dimulai dengan mengirimkan paket SYN berulang ke mesin target sampai *server crash* / malfungsi. Serangan UDP-Lag adalah jenis serangan yang mengganggu koneksi antara *client* dan *server*. Serangan ini sebagian besar digunakan dalam permainan *online* di mana para pemain ingin memperlambat / mengganggu pergerakan pemain lain untuk mengatasi mereka. Serangan ini dapat dilakukan dengan dua cara, yaitu menggunakan sakelar perangkat keras yang dikenal sebagai sakelar jeda atau oleh program perangkat lunak yang berjalan di jaringan dan membanjiri *bandwidth* pengguna lain.



Gambar 2.3 Serangan berbasis Refleksi dan Exploitasi

2.5. Serangan Pada DDoS

Dalam serangan DDoS [3], macam jenis serangan diantaranya akan dijelaskan berikut:

a. ***HTTP Flood***. Dalam jenis serangan ini, penyerang mengeksploitasi permintaan *HTTP GET* atau *POST* untuk menyerang server atau aplikasi. *HTTP Flood* menggunakan *bandwidth* lebih sedikit dan kebanyakan efektif ketika paket permintaan dapat memaksa target untuk mengirim kembali maksimum sumber daya mungkin[3].

b. ***UDP Flood***. Penyerang membanjiri target dengan paket *User Datagram Protocol (UDP)* di port acak dari host acak. Ini memaksa korban untuk terus memeriksa aplikasi mendengarkan pada *port* tersebut tetapi karena tidak ada aplikasi yang ditemukan, ia merespons dengan paket *'Destination Unreachable'* yang menyebabkan habisnya sumber daya[3].

c. ***ICMP Flood*** *(Ping). ICMP Flood* bertujuan membanjiri target dengan permintaan *ICMP (ping)* paket tanpa menunggu balasan[3].

d. ***SYN Flood***. *SYN Flood attack* mengeksploitasi protokol *handshake* tiga arah dari TCP koneksi. Dalam jabat tangan tiga arah, permintaan SYN dijawab oleh SYN-ACK dari tuan rumah dan akhirnya ACK dari pemohon. Penyerang terus menerus mengirim SYN permintaan tanpa menanggapi SYN-ACK korban atau dengan menggunakan alamat IP palsu untuk mengirim permintaan SYN. Bagaimanapun, jabat tangan tetap tidak lengkap dan pada akhirnya menghabiskan lebih banyak sumber daya di korban[3].

e. ***Ping of Death.*** Dalam ping serangan kematian, protokol IP dimanipulasi untuk mengirim malware paket ke target. Ping kematian populer dua dekade lalu tetapi tidak efektif sebagai serangan lain sekarang[3].
f. ***Slowloris***. Serangan *Slowloris* ditujukan ke server web tempat penyerang menggunakan minimal sumber daya untuk menyerang sistem dengan meminta koneksi dengan target dan sesegera mungkin koneksi dibuat, penyerang mencoba untuk menjaga koneksi terbuka selama mungkin dan mengirimkan paket HTTP palsu untuk menguras server web[3].
g. ***NTP Amplification***. Dalam serangan *Amplifikasi NTP*, pelaku menggunakan paket UDP untuk menargetkan server *Network Time Protocol* yang tersedia untuk umum, protokol yang digunakan untuk menyinkronkan jam komputer. Ini adalah serangan amplifikasi karena *query-to respon ratio* adalah serangan seperti itu bisa di mana saja antara 1:20 - 1: 200 atau bahkan lebih[3].
h. ***Zero-day DDoS attacks***. *"Zero-day*" adalah istilah yang digunakan untuk semua serangan yang tidak dikenal atau baru. Serangan-serangan ini mengeksploitasi kerentanan yang belum ada mekanisme pertahanannya[3].

2.6. Weka

Pada penelitian ini bermaksud untuk mengidentifikasi jenis pengklasifikasi terbaik untuk mendeteksi serangan DDOS. Alasan WEKA (Waikato Environment for Knowledge Analysis) *toolkit* digunakan untuk melakukan eksperimen, adalah karena fitur yang memiliki lebih dari 100 metode klasifikasi, yang juga mendukung antarmuka pengguna grafis, dan berbagai alat untuk memvisualisasikan kinerja classifier yang lebih baik, yang juga didukung oleh [9].

Enam kategori pengklasifikasi pembelajaran mesin dari toolkit WEKA dibahas dalam bagian ini. Setiap kategori memiliki beberapa jumlah pengklasifikasi. Kategori tersebut diantaranya *Bayes, Functions, Lazy, Meta, Rules, Trees. Toolkit* WEKA digunakan untuk menganalisis *dataset* dengan algoritma *data mining*. WEKA adalah kumpulan alat klasifikasi data, *regresi*, pengelompokan, aturan asosiasi dan visualisasi. *Toolkit* ini dikembangkan di Jawa dan merupakan

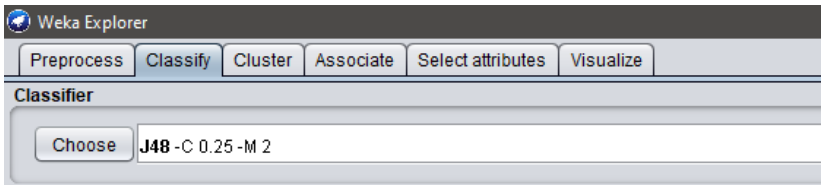
perangkat lunak sumber terbuka yang dikeluarkan di bawah Lisensi Publik Umum GNU.

Didalam WEKA, ada 5 aplikasi di dalamnya yaitu:

a. *Weka Explorer*
b. *Weka Experiment*
c. *Weka Knowledge Flow*
d. *Weka Workbench*
e. *Simple CLI*

Untuk klasifikasi dari *dataset*, *weka explorer* digunakan untuk menghasilkan hasil atau statistik. *Weka Explorer* menggabungkan fitur-fitur berikut di dalamnya:

a. *Preprocess*: Digunakan untuk memproses input data. Untuk keperluan ini *filter* digunakan yang dapat mengubah data dari satu formulir ke formulir lain. Pada dasarnya dua jenis *filter* digunakan yaitu diawasi dan tidak diawasi.
b. *Classify*. Tab Klasifikasi digunakan untuk tujuan klasifikasi. Sejumlah besar pengklasifikasi digunakan dalam weka seperti *Bayes, Functions, Lazy, Meta, Rules, Tree, dan lainnya*.
c. *Cluster*: Digunakan untuk pengelompokan data.
d. *Associate*: Tetapkan aturan asosiasi untuk data.
e. *Select attributes*: Digunakan untuk memilih atribut yang paling relevan dalam data.
f. *Visualize*: Melihat *plot* data 2D interaktif.



Gambar 2.4. Menu pada WEKA

2.7. *Confusion Matrix*

Akurasi adalah suatu hal yang berkaitan dengan kualitas data dan jumlah kesalahan yang terdapat dalam *dataset*. Nilai akurasi akan mengukur sejauh mana informasi pada database digital sesuai dengan nilai sebenarnya atau nilai yang diterima [10].

Perhitungan akurasi didapatkan dari *confusing matrix* yang didapatkan dari nilai TP, TN, FP, FN.

a. *False Positive (FP),* adalah jumlah catatan yang normal tetapi diklasifikasikan sebagai serangan.
b. *False Negative (FN),* adalah jumlah serangan yang benar tetapi diklasifikasikan sebagai catatan yang normal.
c. *True Positive (TP),* adalah jumlah catatan yang serangan dan diklasifikasikan sebagai catatan yang serangan.
d. *True Negative (TN),* adalah jumlah normal yang diklasifikasikan sebagai normal.

Tabel 2.2. *Confusion Matrix*

	Actual Label Positive	*Actual Label Negative*
Predicted Label Positive	TP	FP
Predicted Label Negative	FN	TN

2.8. Cross-Validation

Validasi silang adalah teknik validasi model untuk menilai bagaimana hasil analisis statistik akan digeneralisasi ke kumpulan data independent [11]. Ini terutama digunakan dalam pengaturan di mana tujuannya adalah prediksi, dan orang ingin memperkirakan seberapa akurat model prediksi akan dilakukan dalam praktik (catatan: kinerja = penilaian model).

Namun, validasi silang dapat digunakan untuk membandingkan kinerja spesifikasi pemodelan yang berbeda (yaitu model dengan dan tanpa interaksi, dimasukkannya pengecualian istilah *polinomial,* jumlah simpul dengan *splines* kubik terbatas, dll). Selain itu, validasi silang dapat digunakan dalam pemilihan variabel dan pilih tingkat fleksibilitas yang sesuai dalam *model* (catatan: fleksibilitas: = pemilihan model). Penilaian *Model* digunakan untuk membandingkan kinerja berbagai spesifikasi pemodelan. Pemilihan *Model* digunakan untuk memilih tingkat fleksibilitas yang sesuai dalam model.