ISSN 2460-1160

Klasifikasi Jenis serangan DOS dan Probing pada IDS menggunakan metode K- *Nearest Neighbor*

Yuri Ariyanto[1], Vipkas Al Hadid Firdaus[2], Hanifa Pramana[3]

[1,2,3] Teknik Informatika, Teknologi Informasi, Politeknik Negeri Malang
[1] yuri@polinema.ac.id, [2] vipkas@polinema.ac.id, [3] hanifapramana17@gmail.com

***Abstrak*— IDS (Intrusion Detection System) adalah sebuah aplikasi perangkat keras atau perangkat lunak yang otomatis bekerja untuk memonitor kejadian pada sebuah jaringan komputer dan sekaligus menganalisais masalah keamanan jaringan. Permasalahan muncul ketika aktifitas mencurigakan tidak pada aturan kemanan yang terdaftar, Hal ini menyebabkan menurunnya fungsi IDS dalam mengklasifikasikan aktifitas jaringan. Pada penelitian ini untuk melindungi jaringan dari ancaman yaitu dengan cara membangun sistem deteksi Intrusion Detection System (IDS) menggunakan metode K-Nearest Neighbor untuk mengklasifikasikan jenis serangan normal, DOS, Probing. Tujuan penelitian ini adalah untuk memonitor traffic atau lalu-lintas data pada sebuah jaringan dan menentukan apakah trafic aman, mencurigakan atau terindikasi serangan. Penelitian ini mengguakan dataset KDDCUP 1999.Berdasarkan hasil penelitian ini berupa uji coba sistem menunjuk kan bahwa parameter nilai K sangat berpengaruh terhadap hasil klasifikasi dan akurasi yang dihasilkan. Rata-rata akurasi cenderung menurun seiring dengan penambahan nilai K sedangkan peningkatan jumlah data training turut disertai dengan peningkatan hasil akurasi. Hasil akurasi tertinggi pada pengujian ini sebesar 90% pada saat jumlah data training 3000 data testing 150 dengan menggunakan nilai K= 5.**

***Kata kunci*—** *K-Nearest Neighbor,Klasifikasi, Intrusion Detection Sytem, KDDCUP*

I. PENDAHULUAN

Keamanan menjadi salah satu teknologi yang perlu diperhatikan ketika suatu sistem yang terkoneksi dengan sistem jaringan komputer menjadi hal yang sangat penting. Pada saat ini kebutuhan manusia sangat tergantung dengan adanya informasi ataupun data, khususnya informasi atau data digital. Semakin besar kebutuhan adanya informasi semakin meningkat pula insiden atau gangguan keamanan terhadap sistem jaringan yang meningkat tajam. Maka dari itu diperlukan keamanan dalam sistem komputer untuk mencegah dari beberapa serangan.

Upaya dalam melindungi jaringan dari ancaman serangan ialah membangun sistem deteksi Intrusion Detection System (IDS) untuk memonitor traffic atau lalu-lintas data pada sebuah jaringan dari ancaman-ancaman intruder (penyerang). Intrusion detection adalah proses memonitor kejadian pada sistem komputer atau jaringan dan menganalisanya untuk memberikan tanda insiden yang mungkin, yang mana yang merupakan pelanggaran atau mendekati pelanggaran sebuah kebijakan keamanan komputer, kebijakan penggunaan yang disetujui atau praktik keamanan standar. IDS berfungsi untuk mengidentifikasi traffic atau lalu-lintas data pada sebuah jaringan komputer dimana IDS dapat menentukan apakah traffic aman, mencurigakan atau bahkan terindikasi merupakan serangan [1].

Dalam penelitian Intrusion Detection System (IDS) yang menggunakan data latih yang diambil dari Knowledge Discovery In Database (KDD) dataset cup 1999 yang dikeluarkan oleh DARPA (Defense Advances Research Project Agancy), Knowledge Discovery In Database (KDD) merupakan istilah lain dari data mining yaitu aktifitas mengumpulkan dan memakai data yang berukuran besar dalam rangka mendapatkan pola atau hubungan data. Keluaran data mining, selanjutnya dapat digunakan untuk pengambilan keputusan di waktu mendatang [2].

Permasalahan muncul ketika terdapat aktifitas-aktifitas mencurigakan atau aktifitas tersebut adalah serangan tetapi tidak terdaftar pada aturan keamanan yang terdaftar, sehingga hal tersebut sangat berbahaya bagi jaringan komputer. Oleh karena itu dibutuhkan sebuah sistem klasifikasi serangan yang berfungsi untuk mengklasifikasi anomali lalu lintas jaringan yang ada. Pada penelitian ini akan melakukan klasifikasi dengan menggunakan dataset KDD CUP 1999 sebagai data training dan data testing dan selanjutnya diproses pengklasifikasian menggunakan metode K-*Nearest Neighbor* dan dari klasifikasi tersebeut dapat mengklasifikasikan kelas-kelas jenis serangan pada dataset KDD CUP 99 dengan akurat.

Dengan adanya penelitian yang berjudul "Implementasi Metode K Nearest Neighbor dalam mengklasifikasi jenis serangan pada Intrusion Detection System" , hasil klasifikasi tersebut juga dapat digunakan menjadi dasar untuk membuat aturan baru yang akan didaftarkan pada aplikasi IDS yang digunakan. peningkatan serangan menyebabkan data yang harus dianalisis menjadi sangat besar, mengakibatkan keterbatasan mendeteksi jenis-jenis serangan. penggunaan data mining merupakan solusi dalam membantu pengelolaan data. Berdasarkan penelitian oleh M. Nikitha dan M.A Jabbar pada tahun 2019 yang berjudul "K Nearest Neighbor Based Model for Intrusion Detection System" menunjukkan bahwa algoritma KNN menjadi algoritma yang baik untuk melakukan klasifikasi setelah dibandingkan dengan metode SVM dan Decision Tree dengan tingkat akurasi sebesar 99,96% sedangkan untuk metode SVM sebesar 99,66% dan

untuk Decision Tree sebesar 99,94%, dari penelitian tesebut maka penulis akan menerapkan metode K Nearest Neighbor untuk mengklasifikasi jenis serangan pada Intrusion Detection System (IDS).

II. LANDASAN TEORI

A. (Intrusion Daetection System) IDS

IDS adalah sebuah aplikasi perangkat keras atau perangkat lunak yang otomatis bekerja untuk memonitor kejadian pada sebuah jaringan komputer dan sekaligus menganalisais masalah keamanan jaringan. Sasaran IDS adalah memonitoring aset jaringan sehingga dapat mendeteksi perilaku yang tidak lazim, kegiatan yang tidak sesuai, searngan atau menghentikan serangan (penyusupan) dan bahkan menyediakan informasi untuk menelusuri penyerang. Pada umumnya IDS terbentuk menjadi dua, yaitu:

1. *Network-Based Intrusion Detection System* (NIDS). NIDS merupakan strategi yang efektif untuk melihat *traffic* masuk keluar ataupun traffic di antara host atau di antara segmen jaringan lokal. NIDS biasanya dikembangkan di depan dan di belakang *firewall* dan VPN *gateway* untuk mengukur keefetifan peranti-peranti keamanan tersebut dan berinteraksi dengan mereka untuk memperkuat keamanan jaringan.
2. *Host-Based Intrusion Detection System* (HIDS). HIDS hanya melakukan pemantauan pada perangkat komputer tertentu dalam jaringan HIDS biasanya akan memantau kejadian seperti kesalahan login berkali-kali dan melakukan pengecekan pada file [3].

B. K-Nearest Neighbor

K-Nearest Neighbor (K-NN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari *query instance* yang baru diklasifikan berdasarkan mayoritas dari kategori pada K-NN. Tujuan dari algoritma ini mengklasifikasikan obyek baru berdasarkan atribut dan *training sample*. *Classifier* tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik *query* akan ditemukan sejumlah obyek atau titik training yang paling dekat dengan titik *query*. Klasifikasi menggunakan voting terbanyak diantara klasifikasi dari K obyek. Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *query instance* yang baru.

Syarat nilai K adalah tidak boleh lebih besar dari jumlah data latih, dan nilai K harus ganjil dan lebih dari satu [9].

Algoritma metode K-NN sangatlah sederhana, bekerja berdasarkan jarak terpendek dariquery instance ke training sample untuk menentukan K-NN-nya. Training sample diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasitraining sample. Sebuah titik pada ruang ini ditandai kelas c jika kelas c merupakan klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat dari titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan Euclidean Distance yang direpresentasikan pada persamaan sebagai berikut [6]:

$$D(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (1)$$

Keterangan :

D : jarak antara titik pada data training x dan titik data testing y yang akan diklasifikasi, dimana x=x1,x2,…,xi dan y=y1,y2,…,yi.
x : data uji.
y : data latih.
i : merepresentasikan nilai atribut.
n : merupakan dimensi atribut.

Rumus mengimplementasikan metode K-NN dapat dilihat pada flowchart gambar berikut:

Gambar 1 Flowchart K-Nearest Neighbor

C. KDD CUP 1999 Dataset

Pada tahun 1999, ACM Special Interest Group on Knowledge Discovery and Data Mining adalah organisasi profesional terkemuka penambang data, menyelenggarakan kompetisi di dunia yang mempertemukan berbagai researcher, akademisi dan praktisi untuk dapat memberikan bantuan menyelesaikan kasus yang diberikan dalam kompetisinya tersebut. Kompetisi tersebut adalah *Knowledge Discovery in Database* (KDD) Cup 99 yang bertema Computer Network Intrusion Detection. Dataset KDD CUP 99 dikeluarkan oleh DARPA (*Defense Anvanced Research Projects Agency*) dan dikelola oleh MIT Loncoln Lbs. Kumpulan dari data ini digunakan sebagai alat kompetisi internasional ilmu pengetahuan dan data mining yang ke 3, yang diadakan secara bersamaan dengan konfrensi internasional ilmu pengetahuan dan data mining KDD-99 yang ke lima, tujuan dari kompetisi adalah untuk membangun detektor intrusi jaringan, yang mampu membuat model perbedaan prediksi antara koneksi "buruk" disebut dengan gangguan atau serangan dan baik disebut koneksi normal [2].

Keuntungan menggunakan dataset ini adalah [3]:

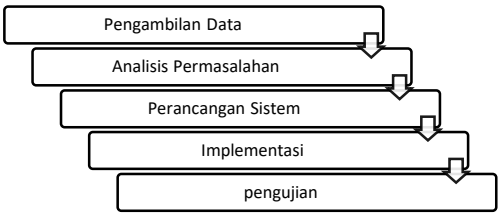
- Tidak ada record yang berlebihan di dalam train set, jadi classifier tidak akan menghasilkan hasil yang bias.
- Tidak ada duplikat record pada test set yang memiliki reduction rates yang lebih baik.
- Jumlah record yang dipilih dari setiap level grup yang berbeda berbanding terbalik dengan persentasi record didalam dataset KDD asli.

Dataset pelatihan KDD ini merupakan data rekam koneksi yang terdiri dari 1 jenis data normal dan 22 jenis data serangan yang dikelompokkan kedalam empat tipe intrusi. Dataset yang diperoleh

dari UCI Repository terdiri dari sekitar 4.900.000 vektor koneksi tunggal yang masing-masing berisi 41 fitur dan diberi label sebagai sebuah serangan atau normal, dengan tipe satu jenis serangan tertentu

III. METODELGI PENELITIAN

Pengimplementasian sistem deteksi serangan meggunakan metode K-Nearest Neighbor, pengembangan aplikasi yang digunakan adalah waterfall, terdapat pada gambar berikut:



Gambar 2 Metode *Waterfall*

A. *Pengambilan Data*

Pengumpulan data dan informasi dengan cara membaca referensi dari buku maupun dari internet terkait IDS, klasifikasi serangan pada IDS, dan *K-Nearest Neighbor*.

Observasi dilakukan dengan mengakses sumber dari http://kdd.ics.uci.edu yang menyediakan data KDD CUP 1999 yang dikeluarkan oleh DARPA (*Defense Advance Research Project Agancy*).

B. *Analisis Permasalahan*

1) Analisis Permasalahan

Adapun penjelasan data yang akan digunakan adalah sebagai berikut:

Atribut/fitur data yang digunakan dalam penelitian ini berjumlah 41 fitur, koneksi dari label normalatau attack, dikategorikan dalam 4 kelas kategori utama seperti pada tabel 1 [7].

TABEL 1 KARAKTERISTIK DASAR DARI INTRUSI PADA DATASET KDD'99

Data Set	Anomaly		Misuse		Normal
	Dos	Probe	U2R	R2L	
10% KDD	391458	4107	52	1126	97277
Corrected KDD	229853	4166	70	16347	60593
Seluruh KDD	3883370	41002	52	1126	972780

2) Analisa Tahapan Knowledge Discovering in Data (KDD)

Tahapan proses pengolahan data pada dataset KDD CUP 99:

a) Data Selection

Proses seleksi pada data penelitian, berupa seleksi fitur atribut yang akan digunakan dan seleksi kelas serangan yang berjumlah 3 kelas yang terbagi yaitu Normal, DoS, dan Probe.

Tahap selection fitur pada penelitian ini mengacu pada penelitian sebelumnya oleh Donny Mongkarang, Noor Akhmad Setiawan, dan Adhistya Erna Permanasri dengan judul "Implementasi Data Mining dengan seleksi Fitur untuk Klasifikasi Serangan Pada Intrusion Detection System (IDS)" ,pada penelitian tersebut menyatakan bahwa hasil seleksi fitur yang dilakukan dengan 10-folds cross validation untuk memproses daftar fitur signifikan dari dataset KDD dan dengan evaluasi teknik seleksi fitur genetic search direduksi menjadi 13 fitur. Dataset hasil seleksi fitur tersebut akan digunakan untuk penelitian ini.

Berikut merupakan tabel seleksi fitur yang akan digunakan.

TABEL 2 SELEKSI FITUR

Fitur	Tipe
Duration	continuous
protocol_type	symbolic
Service	symbolic
Flag	symbolic
src_bytes	continuous
dst_bytes	continuous
Land	symbolic
srv_count	continuous
same_srv_rate	continuous
diff_srv_rate	continuous
dst_host_diff_srv_rate	continuous
dst_host_same_src_port_rate	continuous
dst_host_srv_diff_host_rate	continuous

b) Data Transformation

Proses pada data input yang bertipe text akan terlebih dahulu diubah kedalam bentuk numerik. Selanjutnya transformasi yang akan dilakukan adalah normalisasi.

Tujuan dari normalisasi ini agar data berada pada range (0-1) sehingga sebaran datanya tidak terlalu jauh.

Proses transformasi ini dilakukan berdasarkan penelitian sebelumnya oleh [8]. Berikut ialah hasil proses transformasi pada atribut protocol, service dan flag pada tabel 3.

TABEL 3 TRANSFORMASI PROTOCOL, SERVICE DAN FLAG

Atribut	**Transformasi**	
Protocol	tcp:1 udp:2 icmp:3	
Service	http:1,private:2, domain_u:3, smtp:4, ftp_data:5, eco_i:6, other:7, ecr_i:8, telnet:9, finger:10, ftp:11, auth:12, Z39_50:13, uucp:14, courier:15, bgp:16, whois:17, uucp_path:18, iso_tsap:19, time:20, imap4:21, nnsp:22, vmnet:23, urp_i:24, domain:25, ctf:26, csnet_ns:27, supdup:28, discard:29, http_443:30, daytime:31, gopher:32, efs:33, systat:34, link:35, exec:36,	hostnames:37, name:38, mtp:39, echo:40, klogin:41, login:42, ldap:43, netbios_dgm:44, sunrpc:45, etbios_ssn:46, netstat:47,netbios_ns:48, ssh:49, kshell:50, nntp:51, pop_2:52, sql_net:53, IRC:54, ntp_u:55, rje:56, pop_3:57,remote_job:58, X11:59,printer:60, shell:61, urh_i:62, red_i:63, tim_i:64, pm_dump:65, tftp_u:66, aol:67, harvest:68, http_8001:69, http_2784:70
Flag		S2:8,

	SF:1, S0:2, REJ:3, RSTR:4, RSTD:5, S1:6, SH:7,	

Selanjutnya dilakukan proses normalisasi dengan tujuan mengatasi sebaran data agar nilai dari masing-masing varibel tidak terlalu jauh. Nilai data pada setiap atribut akan diubah pada rentang 0-1. Proses normalisasi ini akan dilakukan pada setiap data penelitian. Berikut proses normalisasi dilakukan dengan dengan rumus normalisasi:

$$v^i = \frac{v - min_a}{max_a - min_a} (new\_max_a - new\_min_a) + new\_min_a \quad (2)$$

Keterangan :

v^i : Data baru setelah normalisasi

v : Data sebelum normalisasi

new_max_a : Batas nilai max baru adalah 1

new_min_a : Batas nilai min baru adalah 0

max_a : Nilai maximum pada kolom

min_a : Nilai minimum pada kolom

c) Klasifikasi K-Nearest Neighbor

Proses menerapkan metode dari data mining untuk mengolah data yang ada. Metode ini melakukan klasifikasi serangan jaringan komputer pada dataset KDD CUP 99 berdasarkan data latih yang jaraknya paling dekat. Untuk mencari jarak tersebut dihitung menggunakan rumus *Euclidean*.

Pada tahapan ini metode *K-Nearest Neighbor* juga menentukan nilai k. Kemudian menentukan kelas mayoritas tertinggi dari nilai k yang ditentukan. Selanjutnya akan menghasilkan hasil klasifikasi untuk menentukan kelas pada data uji. Output yang didapat adalah klasifikasi hasil dari dataset KDD CUP 99.

Penelitian ini akan membagi permasalahan menjadi 3 klasifikasi serangan yaitu DOS (*Denial Of Service)*,Normal (serangan normal), *Probe* (*scanning).*

C. *Perancangan Sistem*

Perancangan desain dan sistem pada klasifikasi jenis serangan pada dataset KDD CUP 99 meliputi: *Use Case Diagram, Activity Diagram, Class Diagram,* dan Rancangan *database* untuk proses gambran yang akan dikerjakan dan mendefiniskan arsitektur sistem secara keseluruhan.

D. *Implementasi*

Langkah-langkah pada tahap implementasi yaitu:

1. Mengunduh File Intrusion detection evaluation dataset.
2. Mengolah data yang tersedia menggunakan metode K-Nearest Neighbor.
3. Membangun sistem berdasarkan perancangan yang telah dibuat sebelumnya.

TABEL 4 RINCIAN DATA TRAINING DAN DATA TESTING

No	Data		RINCIAN
		Jumlah Data			
1	Data Training	1500	400	600	500
		2000	500	800	700
		3000	500	1500	1000
2	Data Testing	150	40	60	50

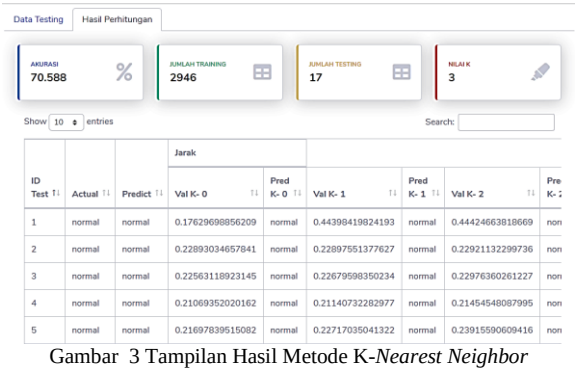
Penelitian menggunakan metode K-*Nearest Neighbor* untuk metode pengkalsifikasian polaserangan baru berdasarkan pola-pola serangan yang sudah ada pada data *training.* Metode K-*Nearest Neighbor* diterjemahkan menggunakan pemrograman PHP yang kemudian hasil dari klasifikasi metode tersebut akan ditampilkan pada *web interface.*

E. *Pengujian*

Pengujian dilakukan dengan blackbox dan pengujian akurasi.

IV. IMPLEMENTASI DAN PENGUJIAN

A. *Tampilan hasil klasifikasi menggunakan metode K-Nearest Neighbor*



Gambar 3 Tampilan Hasil Metode K-*Nearest Neighbor*

Pada gambar 3 ditampilkan halaman hasil perhitungan klasifikasi menggunakan metode K-*Nearest Neighbor*. Ditampilkan perbandingan pelabelan, kolom actual merupakan label yang sebenarnya dan kolom predict adalah hasil pelabelan dari sistem klasifikasi.

B. *Pengujian*

1) Pengaruh Nilai K Terhadap Tingkat Akurasi

Pada metode KNN nilai k menjadi komponen penting yang berpengaruh terhadap akurasi yang akan dilkasifikasi. Pada penilitian ini digunakan rentang nilai k dari 3 sampai 9, sehingga perlu dilakukan pengujian untuk mengetahui berapa nilai k yang menghasilkan tingkat akurasi yang lebih baik. Pengujian dilakukan dengan melakukan percobaan menggunakan data training sebanyak 2000 data dan data testing sebanyak 150 data, diklasifikasikan dengan nilai k dengan nilai ganjil untuk menghindari kesamaan nilai jarak. Rincian data yang akan digunakan antaralain pada tabel 5 berikut

TABEL 5 RINCIAN NILAI K YANG DIGUNAKAN

No	Data Testing	Data Testing	Nilai K
1	2000 data	150 data	3
2	2000 data	150 data	5

3	2000 data	150 data	7
4	2000 data	150 data	9

Setelah itu mendapatkan hasil akurasi masing-masing perbandingan nilai akurasi ditunjukkan pada tabel 6 sebagai berikut

TABEL 6 AKURASI NILAI K

No	Data Training	Data Testing	Nilai K	Akurasi
1	2000	150	3	80%
2	2000	150	5	81,33%
3	2000	150	7	78,67%
4	2000	150	9	78%

Dari tabel diatas, dapat dilihat bahwa besarnya nilai k memiliki akurasi yang lebih baik untuk mengklasifikasi adalah k=5.

2) Pengaruh Jumlah Data Training berbeda dengan Jumlah Data Testing Tetap

Pada pengujian ini menggunakan nilai K yang sudah diuji sebelumnya seperti yang ada pada tabel 6. , pada tabel tersebut nili K= 5 memiliki akurasi lebih baik dibandingkan nilai K yang lain, maka nilai K= 5 yang jadikan acuan untuk menguji pengaruh jumlah data *training* berbeda dengan jumlah data testing tetap. Rincian data yang akan digunakan antaralain pada tabel 7 berikut

TABEL 7 TABEL RINCIAN DATA TRAINING DAN DATA TESTING

No	Data Training	Data Testing	Nilai K
1	1500	150	5
2	2000	150	5
3	3000	150	5

didapatkan hasil akurasi masing-masing, perbandingan nilai akurasi ditunjukkan pada tabel 8 sebagai berikut

TABEL 8 TABEL AKURASI JUMLAH DATA TRAINING TERHADAP DATA TESTING

No	Data Training	Data Testing	Nilai K	Akurasi
1	1500	150	5	78,7%
2	2000	150	5	81,333%
3	3000	150	5	90%

Dari tabel diatas, dapat dilihat peningkatan jumlah data training turut disertai dengan peningkatan hasil akurasi.

V. KESIMPULAN DAN SARAN

A. *Kesimpulan*

Berdasarkan pembahasan dari penelitian ini dapat

ditarik kesimpulan sebagai berikut :

1. Metode KNN dapat digunakan untuk melakukan klasifikasi serangan secara otomatis
2. Akurasi maksimum yang dihasilkan pada penelitian ini sebesar 90% pada saat jumlah data training 3000 data dan akurasi minimum sebesar 78,666% pada saat jumlah data training 1500 data, dan nilai K=5 ,Secara umum penambahan nilai k akan menyebabkan penurunan akurasi.

B. *Saran*

Beserta masalah yang terjadi dapat diberikan saran sebagai berikut:

1. Menggunakan metode lain sehingga dapat dibandingkan tingkat keakuratannya.
2. Menambahkan atau mengganti dengan parameter yang berbeda.
3. Menggunakan metode pattern matching untuk penyesuaian format data dari user ke format dataset KDDCUP 1999.

VI. DAFTAR PUSTAKA

[1] Prasetyo, A., Affandi, L., & Arpandi, D. (2018). Implementasi Metode Naive Bayes Untuk Intrusion Detection System (IDS). *Jurnal Informatika Polinema*. https://doi.org/10.33795/jip.v4i4.220

[2] Anwar, S., Septian, F., & Septiana, R. D. (2019). Klasifikasi Anomali Intrusion Detection System (IDS) Menggunakan Algoritma Naïve Bayes Classifier dan Correlation-Based Feature Selection. Jurnal

[3] Wirawan, I. N. T., & Eksistyanto, I. (2015). Penerapan Naive Bayes Pada Intrusion Detection System Dengan Diskritisasi Variabel. *JUTI: Jurnal Ilmiah Teknologi Informasi*. https://doi.org/10.12962/j24068535.v13i2.a487

[4] Rivki, M., & Bachtiar, A. M. (2017). Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia. *Jurnal Sistem Informasi*. https://doi.org/10.21609/jsi.v13i1.500

[5] Kartika, J. I., Santoso, E., & Sutrisno. (2017). Penentuan Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor dan Weighted Product (Studi Kasus: SMP Negeri 3 Mejayan). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*.

[6] Arianto, N. (2017). Penerapan Seleksi Fitur Fast Correlation Based Filter Pada Metode Modified K-Nearest Neighbor Untuk Mendeteksi Serangan Pada Jaringan Komputer. *Sistem Seleksi Proposal Tugas Akhir*, *2*(5).

[7] Mongkareng, D., Setiawan, N. A., & Permanasari, A. E. (2017). Implementasi Data Mining dengan Seleksi Fitur untuk Klasifikasi Serangan pada Intrusion Detection System (IDS). *Citee*.

[8] Ibrahim, L. M., Taha, D. B., & Mahmod, M. S. (2013). A comparison study for intrusion database (KDD99, NSL-KDD) based on self organization map (SOM) artificial neural network. *Journal of Engineering Science and Technology*.