e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.5 No.1, Juni 2021
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

KLASIFIKASI TINGKAT KEPARAHAN SERANGAN JARINGAN KOMPUTER DENGAN METODE MACHINE LEARNING

Okki Setyawan[1], Angge Firizkiansah[2], Ahmad Nuryanto[3]
Ilmu Komputer[1], Ilmu Komputer[2], Ilmu Komputer[3]
Universitas Nusa Mandiri[1], Universitas Nusa Mandiri[2], Universitas Nusa Mandiri[3]
14002429@nusamandiri.ac.id[1], 14002408@nusamandiri.ac.id[2],
14002431@nusamandiri.ac.id[3]

Received: April 23, 2021**. Revised:** May 05, 2021**. Accepted:** May 22, 2021**.**
Published: June 20, 2021**. Issue Period:** Vol.5 No.1 **(**2021), Pp.128-133

Abstrak: Jaringan komputer berkembang sangat pesat saat ini, hingga banyak beberapa perangkat elektronik terhubung dengan internet , namun system keamanan yang diadopsi oleh perangkat tersebut haruslah mumpuni agar tidak mudah terserang ancaman dan bahaya. Peneliti ingin mengetahui tingkat keparahan ancaman dari suatu serangan yang dideteksi oleh firewall menggunakan record data dari suatu perusahaan, dengan menggunakan machine learning yaitu K-Nearest Neighbours dan Decission Tree. Pengelompokan tingkat keparahan pada sistem keamanan jaringan komputer biasa disebut severity. Pada penelitian ini pembatasan tingkatan keparahan serangan menjadi 3 bagian dari yang tingkatan paling tinggi yaitu critical, high dan medium. Dataset yang diolah merupakan hasil log pada firewall sebanyak 5999 dengan 23 kolom atau fitur. Yang terbaik dari ketiga metode tersebut diantaranya K-Nearest Neighbours mendapatkan hasil akurasi sebesar 100%, kemudian Decission Tree mendapatkan hasil akurasi sebesar 100%. Dengan hasil pengolahan data tersebut maka metode machine learning sangat cocok digunakan untuk mengklasifikasikan tingkat keparahan serangan jaringan komputer.

Kata kunci: Klasifikasi, Machine Learning, Jaringan.

Abstract: *Computer networks are currently developing very rapidly, so that many electronic devices are connected to the internet, but the security system adopted by these devices must be qualified so they are not vulnerable to threats and dangers. Researchers want to find out how severe the threat of an attack is detected by a firewall using data records from a company, using machine learning, namely K-Nearest Neighbors, Decission Tree. Classification of the severity of a computer network security system is usually called the severity level. In this study, the limitation of the seriousness level of the attack was divided into 3 parts from the highest level, namely critical, high and medium. The processed dataset is logging into the firewall as many as 5999 with 23 columns or features. The best of the three methods are K-Nearest Neighbors getting 100% accuracy and Decission Tree getting 100% accuracy . With the results of this data processing, the machine learning method is very suitable to be used to classify the severity of computer network attacks.*

Keywords: *Classification, Machine Learning, Network*

DOI: 10.52362/jisicom.v5i1.443

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.5 No.1, Juni 2021
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

I. PENDAHULUAN

Jaringan Komputer berkembang dengan sangat pesat, baik di instansi-instansi komersil, dunia akademik, bahkan rumah-rumah penduduk yang membutuhkan akses internet. Internet diakses oleh banyak orang tanpa terkecuali hacker dan cracker. Dengan alasan tertentu mereka melakukan penyusupan yang dapat merugikan para pemilik server dan jaringan komputer. Mereka menggunakan berbagai macam serangan jaringan komputer dengan tools yang dibuat secara mandiri ataupun yang telah ada di pasar. Kecanggihan serangan dan tools pada jaringan komputer berbanding terbalik dengan pengetahuan tentang penyusupan pada jaringan komputer. Dari tahun 1980an hingga tahun 1990an dimulainya penebakan password, mengetahui password, eksploitasi pengetahuan kerentanan, menonaktifkan audit, pencurian, sampai pada sesi pembajakan. Tahun 2000an diawali dengan serangan *Denial of Service* hingga tahun 2014an berkembang menjadi serangan *Distributed Denial of Service* dan enkripsi biner. Efek utama dari serangan jaringan komputer berupa lambatnya akses internet. Selain itu untuk jenis serangan jaringan yang sangat berbahaya dapat mengakibatkan rusaknya data pada server, sehingga hal ini sangat merugikan pengguna ataupun end user yang sedang mengakses. Kegiatan merusak, mengganggu, mencuri data, dan segala hal yang merugikan pemilik server pada jaringan komputer adalah suatu tindak ilegal dan dapat dijatuhkan sangsi secara hukum di pengadilan. Memerangi kejahatan internet telah menjadi porsi utama bagi agen-agen penegak hukum dan intelijen, baik nasional maupun internasional, tanpa kecuali para praktisi bisnis, sampai kepada para pelanggan, dan end user. Umumnya kejahatan internet dimulai dengan mengeksploitasi host-host dan jaringan komputer sehingga para penyusup datang melintasi jaringan, terutama jaringan yang berbasis *TCP/IP* [1]. Salah satu tools yang dapat digunakan untuk melakukan tindakan preventif terhadap serangan jaringan komputer dengan menggunakan firewall. Firewall merupakan perangkat lunak atau sistem perangkat keras yang mengisi lalu lintas jaringan sesuai dengan aturan yang telah ditentukan sebelumnya. Oleh karena itu, firewall ditanamkan pada sistem operasi atau pusat dari arsitektur jaringan yang digunakan. Berdasarkan hal tersebut, firewall akan menganalisa apa yang terjadi dan menerapkan tindakan dengan mengacu pada aturan yang yang telah ditetapkan sebelumnya. Firewall perlu ditempatkan di perimeter eksternal tetapi juga di dalam jaringan untuk segmentasi data yang aman. Penerapan internal firewall adalah praktik terbaik yang relatif baru. Ini sebagian besar didorong oleh kecenderungan bahwa kita tidak dapat lagi membedakan batas jaringan yang nyata dan andal antara lalu lintas jaringan internal tepercaya dan lalu lintas jaringan eksternal yang tidak tepercaya. Seseorang harus mempertimbangkan kemungkinan masalah keamanan yang timbul dalam jaringan tepercaya juga. Misalnya, firewall yang memisahkan dua departemen perusahaan dapat memperlambat atau bahkan menghentikan penyebaran spam, virus, atau ancaman lainnya [2]. *Machine Learning* adalah cabang aplikasi dari *Artificial Intelligence* (kecerdasan buatan) yang fokus pada pengembangan sebuah sistem yang mampu belajar sendiri tanpa harus berulang kali program oleh manusia. Model yang dipakai pada penelitian ini adalah *K-Nearest Neighbours,* dan *Decision Tree* untuk mengklasifikasikan tingkat keparahan serangan jaringan komputer. Pengklasifikasian tingkat keparahan pada sistem keamanan jaringan biasa disebut *severity* atau keparahan. Pada penelitian ini pembatasan tingkatan keparahan serangan menjadi 3 bagian dari yang tingkatan paling tinggi (parah) yaitu *critical*, *high* dan *medium*.

II. METODE DAN MATERI

Penelitian ini dibagi menjadi 2 tahap, tahap pertama studi literatur yang menghasilkan sebuah paper, tahap kedua adalah pemodelan berisi tentang Dataset Analysis (menganalisa data yang terkumpul), Data Preparation, Data Spliting (training dan testing) dan modeling.

2.1. Dataset Analysis

Penelitian ini mengolah data set dari pengambilan data log pada firewall dalam bentuk format csv *(comma separated value)* dengan jumlah data 5999 dan 23 kolom ukuran 290 KB

	attack	severity	crlevel	time	dstport	srcport	bid_n	craction_n	crscore	epid
0	TCP.Split.Han	3.0	medium	16.0	16729.0	61417.0	3435363.0	16384.0	10.0	92565.0
1	TCP.Split.Han	3.0	medium	14.0	7469.0	61059.0	3400097.0	16384.0	10.0	94137.0
2	TCP.Split.Han	3.0	medium	12.0	7469.0	58347.0	3396061.0	16384.0	10.0	94137.0
3	MS.SMB.Server	3.0	medium	11.0	445.0	50770.0	3395607.0	16384.0	10.0	90966.0
4	MS.SMB.Server	3.0	medium	11.0	445.0	50732.0	3395607.0	16384.0	10.0	90966.0

DOI: 10.5236

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.5 No.1, Juni 2021
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

Gambar 1. Dataset

2.2. Data Preparation

a) Data Cleaning

Dari dataset diatas untuk tahap selanjutnya peneliti menghapus data yang tidak digunakan. Adapun yang harus dihapus adalah kolom yang bukan menjadi parameter yang berorientasi pada hasil yang diinginkan. Setelah penghapusan kolom yang tidak diperlukan juga ada data bernilai NAN di masing-masing kolom yang harus dibersihkan atau dihapus supaya data bisa dinilai dengan valid. Dari pembersihan data tersebut dari data awal sebanyak 6466 baris dan 23 kolom menghasilkan data yang siap diolah yaitu menjadi 5999 baris dan 10 kolom.

	attack	severity	crlevel	time	dstport	srcport	bid_n	craction_n	crscore	epid
0	TCP.Split.Han	3.0	medium	16.0	16729.0	61417.0	3435363.0	16384.0	10.0	92565.0
1	TCP.Split.Han	3.0	medium	14.0	7469.0	61059.0	3400097.0	16384.0	10.0	94137.0
2	TCP.Split.Han	3.0	medium	12.0	7469.0	58347.0	3396061.0	16384.0	10.0	94137.0
3	MS.SMB.Server	3.0	medium	11.0	445.0	50770.0	3395607.0	16384.0	10.0	90966.0
4	MS.SMB.Server	3.0	medium	11.0	445.0	50732.0	3395607.0	16384.0	10.0	90966.0
...	...	...	...	...	...	...	...	...	...	...
5994	MS.SMB.Server	3.0	medium	10.0	445.0	56956.0	3391947.0	16384.0	10.0	90966.0
5995	Backdoor.Doub	1.0	critical	10.0	445.0	54088.0	3391936.0	4096.0	50.0	90966.0
5996	MS.SMB.Server	3.0	medium	10.0	445.0	53930.0	3391936.0	16384.0	10.0	90966.0
5997	MS.SMB.Server	3.0	medium	10.0	445.0	53587.0	3391930.0	16384.0	10.0	90966.0
5998	Backdoor.Doub	1.0	critical	10.0	445.0	52367.0	3391930.0	4096.0	50.0	90966.0

5999 rows × 10 columns

Gambar 2. Data Cleaning

b) Data Transformation

Tahap selanjutnya adalah melakukan pengkodean secara otomatis dengan data encoder pada python 3.0. Fitur-fitur yang dilakukan encoder adalah severity, attack, crlevel, dstport, crscore.

DOI: 10.52362/jisicom.v5i1.443

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.5 No.1, Juni 2021
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

severity_n	attack_n	crlevel_n	dstport_n	crscore_n
2	7	2	4	0
2	7	2	3	0
2	7	2	3	0
2	3	2	1	0
2	3	2	1	0
...	...	...	...	...
2	3	2	1	0
0	1	0	1	2
2	3	2	1	0
2	3	2	1	0
0	1	0	1	2

Gambar 3. Data Transformation

c) Feature (X) dan Prediction (Y)

```
fitur = ['attack_n','crlevel_n','dstport_n','crscore_n']
X = data [fitur]
y = data ['severity']
y = y.values.reshape(-1,1)
```

Gambar 4. Fitur training dan testing

2.3. Data Splitting(training dan testing)

Pengukuran dari penilitian ini adalah dengan data splitting yaitu data training sebesar 80% dan sisanya 20% sebagai data testing.

2.4. K-Nearest Neighbours

Metode klasifikasi *K-Nearest Neighbour* (K-NN) merupakan salah satu metode dari algoritma yang paling sederhana yang menemukan data yang tidak teridentifikasi menggunakan *data points* yang diketahui (*nearest neighbor*) dan mengklasifikasikan data dengan ssistem voting [1]. Jika ditambahkan sebuah data yang tidak diketahui labelnya, K-NN mengklasifikasikan data tersebut dengan menghitung jarak terdekat *majority* dengan data yang sudah ada. Untuk menghitung jarak terdekat tersebut, digunakan banyak metode, namun yang paling banyak digunakan adalah *Euclidean distance*.

Karena sangat sederhana, metode ini telah banyak digunakan pada berbagai bidang, seperti *Pattern recognition*, *Image databases*, *Internet marketing*, *Cluster analysis* dan lainnya [1]. Agar tidak terjadi *voting* seri pada *binary classification*, maka diusahakan nilai k berupa bilangan ganjil. Berikut adalah *pseudocode* dari metode *K-Nearest Neighbor*.

DOI: 10.52362/jisicom.v5i1.443

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.5 No.1, Juni 2021
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

```
K <- jumlah dari nearest neighbors
foreach object X in test set do
hitung jarak D(X,Y) antara X dan setiap object Y
neighborhood <- k neighbors pada data training yang
terdekat dengan X
X class <- SelectClass(neighborhood)
    Endfor
```

Gambar 5. Data Transformation

2.5. Decission Tree

Pohon keputusan atau dikenal dengan Decission Tree adalah salah satu metode klasifikasi yang menggunakan representasi struktur pohon (tree) dimana setiap node mempresentasikan nilai dari atribut, dan daun merepresentasikan kelas. Node yang paling atas dari *decision tree* disebut sebagai *root*.

III. PEMBAHASAN DAN HASIL

Dari berbagai pendekatan machine learning yang kami telah pilih pada bab sebelumnya yaitu K-Nearest Neighbors dan Decision Tree. bersumber dari dataset yang telah kami lakukan proses data preparation, kami telah mendapatkan hasil sebagai berikut.

3.1. K-Nearest Neighbour (KNN)

Pengolahan menggunakan metode KNN dari dataset ini menghasilkan hasil sangat baik.

Tabel 1. Hasil K-Nearest Neighbors

No.	KNN	Hasil
1.	**1**	**100 %**
2.	**3**	**100 %**
3.	**5**	**100 %**
4.	**7**	**99 %**
5.	**9**	**99%**

3.2. Decission Tree

Tidak kalah dengan KNN, pengolahan menggunakan model machine learning decision tree dengan berbagai optimasi dari dataset ini menghasilkan hasil sangat baik juga.

Tabel 2. Hasil Decision Tree

No.	Decision Tree (Pre-Pruning)	Hasil
1.	**0**	**100%**
2.	**5**	**100%**
3.	**10**	**100%**
4.	**15**	**100%**
5.	**20**	**100%**

DOI: 10.52362/jisicom.v5i1.443

e-ISSN : 2597-3673 (Online) , p-ISSN : 2579-5201 (Printed)
Vol.5 No.1, Juni 2021
Journal of Information System, Informatics and Computing
Website/URL: http://journal.stmikjayakarta.ac.id/index.php/jisicom
Email: jisicom@stmikjayakarta.ac.id , jisicom2017@gmail.com

Dari hasil pengolahan data yang telah dilakukan didapatlah tingkat hasil akurasi sangat baik dalam menentukan tingkat keparahan serangan jaringan komputer dengan 3 metode *machine learning*.

Tabel 3. Hasil Pengolahan Machine Learning

No.	Machine Learning	Hasil
1.	**KNN**	**100%**
2.	**Decision Tree**	**100%**

IV. KESIMPULAN

Metode *K-Nearest Neighbors* sebesar 100%, serta *Decision Tree sebesar* 100% . Dapat diambil kesimpulan bahwa dari hasil penelitian ini dengan menggunakan metode *machine learning K-Nearest Neighbors* (KNN) dan *Decision Tree*, sehingga ketiga metode *machine learning* tersebut mampu optimal digunakan untuk melakukan pengelompokan atau mengklasifikasikan tingkat keparahan (*severity*) serangan jaringan komputer yang diharapkan oleh peneliti.

REFERENSI

[1] Fadlil, A., Riadi, I., Aji, S., & Dahlan, U. A. (2017). 5665-15020-1-Pb, 3(1).
[2] Valentin, K., & Maly, M. (2013). NETWORK FIREWALL USING ARTIFICIAL Kristi ´ an Valent ´ ın , Michal Mal y, 32, 1312–1327 http://www.ieee.org. [Accessed: 10 Sept. 2010].
[3] H. A. Nimr, "Defuzzification of the outputs of fuzzy controllers," presented at 5th Int. Conf. on Fuzzy Systems, 1996, Cairo, Egypt. 1996.
[4] T. J. van Weert and R. K. Munro, Eds., *Informatics and the Digital Society: Social, ethical and cognitive issues*: IFIP TC3/WG3.1&3.2 Open Conf. on Social, Ethical and Cognitive Issues of Informatics and ICT, July 22-26, 2002, Dortmund, Germany. Boston: Kluwer Academic, 2003.
[5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.
[6] European Telecommunications Standards Institute, "Digital Video Broadcasting (DVB): Implementation guidelines for DVB terrestrial services; transmission aspects," *European Telecommunications Standards Institute*, ETSI TR-101-190, 1997. [Online]. Available: http://www.etsi.org. [Accessed: Aug. 17, 1998].
[7] "A 'layman's' explanation of Ultra Narrow Band technology," Oct. 3, 2003. [Online]. Available: http://www.vmsk.org/Layman.pdf. [Accessed: Dec. 3, 2003].
[8] G. Sussman, "Home page - Dr. Gerald Sussman," July 2002. [Online]. Available: http://www.comm.pdx.edu/faculty/Sussman/sussmanpage.htm. [Accessed: Sept. 12, 2004].
[9] *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
[10] A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

DOI: 10.52362/jisicom.v5i1.443