COSTING: Journal of Economic, Business and Accounting
Volume 7 Nomor 5, Tahun 2024
e-ISSN : 2597-5234

DETECTION OF FALSE POSITIVE ACCURACY IN INTRUSION DETECTION SYSTEMS USING KNN (K NEAREST NEIGHBOR)

DETEKSI AKURASI FALSE POSITIVE PADA SISTEM DETEKSI INTRUSI MENGGUNAKAN KNN (K NEAREST NEIGHBOR)

Ricky Putra Nugraha[1], Benfano Soewito[2]
Binus University[1,2]
ricky.nugraha@binus.ac.id[1], b.soewito@binus.edu[2]

ABSTRACT

Currently there are many attacks that cause attacks or malware to victims on computers, servers, networks, etc. To prevent all of this, a system is needed that can prevent or detect if such an attack occurs by the name of intrusion detection system (IDS). To prevent all this, a system is needed that can prevent or detect if such an attack occurs with the name intrusion detection system (IDS). IDS can detect recognized or unrecognized attacks or malware. However, one of the problems in implementing an intrusion detection system is false positives. False positives can be very dangerous if not handled properly because they can allow attacks to occur undetected. Therefore, in this research, the author reduces false positives using machine learning and compares the performance of KNN machine learning with KNN Genetic Algorithm to determine which algorithm is the best. In this study, researchers used several stages to improve data quality and the level of accuracy obtained, such as: Missing Value handling, Data Transformation, Data Normalization and Euclidean Distance calculation.

Keywords: *K-NN (K Nearest Neighbor), Intrusion Detection System, Genetich Algorithm, Machine Learning*

ABSTRAK

Pada saat ini terdapat banyak serangan yang menyebabkan serangan atau malware kepada korban pada komputer, server, jaringan, dll. Untuk mencegah semua ini, diperlukan sistem yang dapat mencegah atau mendeteksi jika serangan seperti itu terjadi dengan nama intrusion detection system (IDS). IDS dapat mendeteksi serangan atau malware yang dikenali atau tidak dikenali. Namun, salah satu masalah dalam mengimplementasikan intrusion detection system adalah false positive. False positive dapat sangat berbahaya jika tidak ditangani dengan baik karena dapat memungkinkan serangan terjadi tanpa terdeteksi. Oleh karena itu, dalam penelitian ini, penulis untuk mengurangi false positive menggunakan machine learning serta membandingkan performa machine learning KNN dengan KNN Genetika Algoritma untuk menentukan algoritma mana yang terbaik. Dalam penelitian ini, peneliti menggunakan beberapa tahap untuk meningkatkan kualitas data dan tingkat akurasi yang diperoleh, seperti: penanganan Missing Value, Transformasi Data, Normalisasi Data dan perhitungan Euclidean Distance.

Kata Kunci: K-NN (K Nearest Neighbor), Intrusion Detection System, Genetich Algorithm, Machine Learning

PENDAHULUAN

Seiring dengan kemajuan teknologi saat ini maka kejahatan dalam internet meningkat, di mana teknologi informasi memainkan peran krusial dalam hampir setiap aspek kehidupan kita, perlindungan terhadap data dan sistem yang sensitif menjadi semakin penting. Ancaman terhadap keamanan siber terus berkembang dan menjadi lebih kompleks, sehingga diperlukan pendekatan yang lebih canggih untuk mengidentifikasi dan melindungi sistem dari serangan yang berpotensi merusak. Salah satu alat yang memiliki peran sentral dalam upaya perlindungan ini adalah intrusion detection system (IDS), intrusion detection system merupakan komponen penting dalam lanskap keamanan siber yang dirancang untuk mengawasi lalu lintas jaringan dan sistem komputer guna mendeteksi aktivitas yang mencurigakan atau serangan yang dapat mengancam integritas, kerahasiaan, dan ketersediaan data.

IDS memiliki bentuk yaitu IDS Signature: IDS Signature menggunakan database serangan yang sudah dikenal untuk mengidentifikasi serangan pada jaringan. Metode ini menggunakan pola serangan yang sudah dikenal dan membandingkannya dengan pola yang terdeteksi pada jaringan. IDS Signature lebih cocok untuk mengidentifikasi serangan yang sudah dikenal sebelumnya. :

Kelebihan:

- Lebih mudah untuk diterapkan dan diimplementasikan karena tidak memerlukan pemrosesan data yang kompleks.
- Lebih akurat dalam mendeteksi serangan yang sudah dikenal sebelumnya.

Kekurangan:

- Tidak efektif dalam mendeteksi serangan yang tidak diketahui atau serangan yang menggunakan teknik baru.
- Rentan terhadap false positives atau serangan yang sedikit dimodifikasi dari serangan yang sudah dikenal sebelumnya.

IDS ini dapat memonitor lalu lintas masuk ke dalam sistem known atau unknown dengan indikasi false positive atau false negative, sehingga mencegah aktivitas berbahaya. Dalam penelitian ini, permasalahan dalam IDS ini adalah untuk meningkatkan akurasi false positive, dapat diatasi dengan menggunakan machine learning dan penggunaan machine learning untuk sistem deteksi intrusi ini telah dilakukan sebelumnya.

(Balakrishnan Senthilnayaki, 2019) menggunakan Intrusion Detection System untuk mendeteksi dan mencegah serangan pada jaringan dan basis data, namun, masalah dimensionalitas menjadi masalah utama. Untuk mengatasi masalah ini, para peneliti menggunakan metode Seleksi Fitur Fuzzy Rough Set dalam Intrusion Detection System dan memodifikasi KNN. Hasil penelitian menunjukkan nilai (Probe 99,51%), (DoS 99,29%), dan (Lainnya 71,62%).

(Hafsa Benaddi, 2018) Untuk memonitor serangan dalam Intrusion Detection System, para peneliti menggunakan metode algoritma klasifikasi KNN untuk seleksi fitur dan kombinasi PCA Fuzzy Clustering KNN untuk menganalisis komponen utama. Hasil penelitian menunjukkan nilai akurasi sebesar 94,23% untuk DOS, 69,87% untuk R2L, 80,09% untuk U2L, dan 78,86% untuk PROBE.

Penulis akan menggunakan metode machine learning KNN tidak hanya KNN penulis juga akan melakukan kombinasi antara KNN dengan Genetika Algoritma kemudian membandingkan tingkat akurasi serta parameter lain seperti Precision Recall dan ROC sehingga dapat diketahui metode machine learning apa yang dapat menghasilkan nilai akurasi terbaik, Selanjutnya untuk penggunaan dataset penulis akan menggunakan NSL KDD Dataset atau Network Security Lab Dataset (https://www.kaggle.com/datasets/hassan06/nslkdd/code?resource=download) sebagai dataset, yang merupakan versi terbaru dari KDD Cup 1999. Awalnya diselenggarakan oleh ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD), kompetisi ini berfokus pada masalah umum dalam mendeteksi serangan jaringan yang terus terjadi hingga saat ini, yang menarik minat penulis untuk menggunakannya. Dataset KDD terdiri dari 41 atribut dan 5 kelas, namun penulis hanya akan fokus pada 29 atribut yang menjadi fokus penelitian ini kemudian juga berdasarkan penelitian sebelumnya maka untuk

dapat meningkatkan nilai akurasi yang didapat maka penulis akan menggunakan parameter tuning dimana melakukan parameter tuning diharapkan mampu meningkatkan tingkat akurasi pada KDD Dataset.

kerangka kerja

A. Machine Learning

Machine learning adalah cabang ilmu komputer yang berfokus pada pengembangan algoritma dan teknik yang memungkinkan mesin atau komputer untuk belajar dari data dan meningkatkan kinerjanya dalam tugas-tugas tertentu saat menerima pengalaman dan informasi baru. Dalam machine learning, model komputer dirancang untuk melakukan tugas-tugas tertentu seperti klasifikasi, prediksi, pengenalan pola, dan pengelompokan, menggunakan data yang disediakan. Algoritma machine learning digunakan untuk mempelajari pola dan hubungan dalam data, dan membangun model yang dapat digunakan untuk memprediksi hasil yang diinginkan dari data baru. (Wenchao Li, Ping Yi, Yue Wu, 2014) Peneliti mengusulkan penggunaan Intrusion Detection System dengan algoritma klasifikasi KNN pada Jaringan Sensor Nirkabel (Wireless Sensor Network). S

istem ini diimplementasikan menggunakan GAINZZ Zig-bee Nodes dan dapat mencapai efisiensi tinggi dan deteksi intrusi yang cepat. Hasil penelitian menunjukkan bahwa pada Eksperimen 1, ketika nilai K tidak berada di antara jumlah node serangan dan node normal, maka akan mempertimbangkan tingkat alarm palsu. Eksperimen 2 dan 3 menunjukkan deteksi yang jelas dan deteksi yang tertunda ketika nilai K disesuaikan. Eksperimen 4, 5, dan 6 tidak mendeteksi node serangan karena nilai cutoff yang tidak sesuai, tetapi hanya frekuensi serangan yang berbeda.

B. K-NN (K-Nearest Neighbor)

K-Nearest Neighbor (K-NN) adalah metode yang menggunakan algoritma berbimbing di mana hasil sampel baru diklasifikasikan berdasarkan mayoritas kategori pada K-NN. Tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan sampel latihan, tanpa menggunakan model yang harus disesuaikan, dan hanya mengandalkan ingatan (memori). Diberikan sebuah titik uji, K objek (titik pelatihan) terdekat dengan titik uji akan diidentifikasi. Klasifikasi kemudian dilakukan dengan voting mayoritas di antara K objek tersebut. Algoritma K-NN menggunakan klasifikasi tetangga terdekat sebagai nilai prediksi untuk sampel uji baru.

Penentuan nilai jarak metrik akan menggunakan Euclidean Distance, yaitu untuk menentukan nilai antara dua titik.

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (xi - yi)^2} \quad (1)$$

Data akan dinormalisasi menggunakan metode MinMax scalar, di mana data yang telah diubah akan diubah skala menjadi rentang 0-1 secara matematis. Rumus MinMax scalar adalah sebagai berikut: misalkan X adalah sebuah set data, dan xi adalah data ke-i.

$$xi\,transform = \frac{(xi - min(x))}{(max(x) - min(x))} \quad (2)$$

C. *Genetic Algorithm*

Genetic Algorithm atau Genetika Algoritma (GA) adalah teknik optimisasi komputasional yang terinspirasi oleh prinsip seleksi alam dan genetika. GA digunakan untuk

mengeksplorasi ruang solusi dari suatu masalah dengan menggunakan teknik seperti mutasi, rekombinasi, dan seleksi, dengan tujuan untuk menemukan solusi optimal atau perkiraan solusi optimal.

Pada dasarnya, GA mewakili solusi kandidat sebagai kromosom atau individu, di mana setiap kromosom terdiri dari kumpulan parameter yang mendefinisikan solusi potensial untuk masalah yang dihadapi. Kromosom-kromosom ini kemudian digabungkan dan dimutasi untuk menghasilkan populasi solusi baru, yang dievaluasi menggunakan fungsi kecocokan (fitness function) untuk menentukan seberapa baik mereka memecahkan masalah. Individu-individu yang lebih cocok dipilih sebagai orangtua untuk generasi berikutnya, di mana mereka mengalami crossover dan mutasi untuk menghasilkan keturunan baru. Proses ini diulang untuk beberapa generasi dengan harapan bahwa setiap generasi akan semakin mendekati solusi optimal.

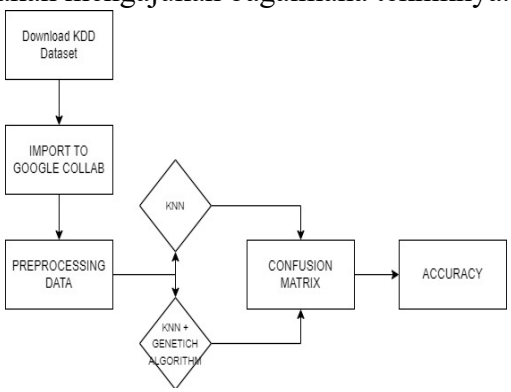
Selain itu, GA dikenal karena kemampuannya dalam mengatasi data noise, tidak lengkap, atau tidak pasti, serta fleksibilitasnya dalam menggabungkan pengetahuan atau batasan khusus dalam domain masalah. Namun, meskipun GA tidak menjamin menemukan solusi optimal, ia dapat mengalami konvergensi prematur atau kompleksitas komputasi.

D. Data Mining

Data mining adalah teknik atau metode pengolahan data yang digunakan untuk mengekstraksi informasi yang berguna atau pola tersembunyi dari kumpulan data besar. Tujuannya adalah untuk menemukan pola, hubungan, atau informasi yang berguna yang dapat digunakan untuk meningkatkan pengambilan keputusan bisnis dan menyelesaikan masalah kompleks. Proses data mining melibatkan beberapa tahapan, seperti pra-pemrosesan data, pemilihan atribut, pemodelan data, dan validasi. Beberapa teknik data mining yang paling umum digunakan termasuk analisis cluster, analisis asosiasi, dan klasifikasi.

METODE PENELITIAN

Pada bagian ini penulis mengusulkan perbandingan dari dua algoritma yaitu KNN dengan KNN Genetika Algoritma untuk untuk mencari tingkat akurasi terbaik dari atribut dataset KDD yang akan diklasifikasikan, terlebih dahulu. Kami akan menjelaskan apa itu KNN dan Algoritma Genetika, kemudian kami akan mengajukan bagaimana tekniknya.



Gambar.1 Rancangan Sistem

a. Download KDD Dataset

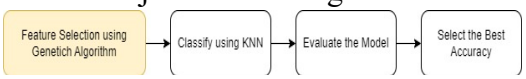
Dataset NSL-KDD (NSL-KDD: NSL-KDD Network Intrusion Detection Data Set) adalah salah satu dataset yang umum digunakan dalam penelitian deteksi intrusi jaringan. Dataset NSL-KDD adalah versi terbaru dari Dataset KDD Cup 1999, yang telah banyak digunakan sebagai dataset standar untuk mengevaluasi Intrusion Detection System pada jaringan.

Dataset NSL-KDD memiliki beberapa perbedaan dengan Dataset KDD Cup 1999. Dataset NSL-KDD telah diperbarui dengan menghilangkan beberapa kelemahan dan kekurangan dari Dataset KDD

Cup 1999. Selain itu, dataset NSL-KDD memiliki jumlah sampel yang lebih besar, sekitar 125.000, dibandingkan dengan Dataset KDD Cup 1999 yang hanya memiliki sekitar 4 juta sampel.

b. Import To Google Collab
Tahapan kedua adalah impor ke Google Colab, yang berarti file tersebut siap untuk diteliti dan alat yang akan digunakan untuk penelitian adalah Google Colab. Google Colab adalah alat yang akan membantu penulis dalam mengklasifikasikan apakah suatu atribut merupakan ancaman atau tidak, dan menunjukkan tingkat akurasi yang dihasilkan. Selain itu, Google Colab dapat digunakan dengan query Python.

c. Proses Kerja Genetika Algoritma



Gambar 2. Proses Kerja Genetika Algoritma

Kombinasi antara Genetika Algoritma dengan KNN dapat digunakan untuk meningkatkan kinerja model model KNN untuk menemukan subset fitur yang optimal dengan feature selected dari genetika algoritma, berikut Langkah langkahnya :

1) Inisialisasi individu secara acak.
2) Menetapkan fitness values untuk setiap individu pada populasi.
3) Melakukan individual selections dalam populasi untuk menciptakan generasi baru
4) Melakukan crossover pada individu terpilih.
5) Melakukan mutasi untuk menghindari persamaan generasi dari hasil crossover dan parent population.
6) Mengulangi Langkah sebelumnya sampai kriteria terpenuhi.

d. Proses Kerja Tuning pada KNN
Pada penelitian ini penulis menggunakan proses tuning knn yang dimana menggunakan range antar nilai k dimana nilai k=29 yang digunakan untuk mengklasifikasikan atau melakukan prediksi menggunakan nilai optimal, tujuannya adalah untuk menemukan nilai k yang paling baik dengan menunjukkan hasil metric evaluasi yaitu : Accuracy, Recall dan Precission.

e. Preprocessing Data
Tahapan ketiga adalah pra-pemrosesan data. Pra-pemrosesan data adalah proses persiapan data untuk digunakan di mana data yang akan digunakan akan melalui beberapa tahap, seperti pemilihan data, transformasi data, dan reduksi data. Setelah pra-pemrosesan data selesai, data dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian. Setelah data diproses, dataset dapat dianalisis menggunakan metode machine learning. Tahapan ketiga adalah pra-pemrosesan data. Pra-pemrosesan data adalah proses persiapan data untuk digunakan di mana data yang akan digunakan akan melalui beberapa tahap, seperti pemilihan data, transformasi data, dan reduksi data. Setelah pra-pemrosesan data selesai, data dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian. Setelah data diproses, dataset dapat dianalisis menggunakan metode machine learning.

- Seleksi

Dalam proses ini, dilakukan seleksi data set, menciptakan data set target, atau berfokus pada subset variabel (sampel data) dalam dataset NSL KDD CUP 1999. Dataset ini berisi 41 atribut. Dari 41 atribut tersebut,

akan dilakukan seleksi untuk kemudian menjalani proses transformasi dan klasifikasi, dan penulis hanya menggunakan 29 atribut saja.

Tabel 1. Dataset Yang Digunakan

Num	Attributes	Desc
1	Duration	A1
2	Protocol_type	A2
3	Service	A3
4	Flag	A4
5	Src_bytes	A5
6	Dst_bytes	A6
7	Hot	A7
8	Logged_in	A8
9	Num_compromised	A9
10	Num_root	A10
11	Is_guest_login	A11
12	Count	A12
13	Srv_Count	A13
14	Serror_rate	A14
15	Srv_serror_rate	A15
16	Rerror_rate	A16
17	Srv_rerror_rate	A17
18	Same_srv_rate	A18
19	Diff_srv_rate	A19
20	Srv_diff_host_rate	A20
21	Dst_host_srv_count	A21
22	Dst_host_same_srv_rate	A22
23	Dst_host_diff_srv_rate	A23
24	Dst_host_same_src_port_rate	A24
25	Dst_host_srv_diff_host_rate	A25
26	Dst_host_serror_rate	A26
27	Dst_host_srv_serror_rate	A27
28	Dst_host_rerror_rate	A28
29	Dst_host_srv_rerror_rate	A29

- Missing Value
  Pada tahap missing value, penulis akan melakukan missing value dimana missing value adalah memastikan bahwa dataset yang dimiliki lengkap tidak ada yang kosong atau null, pada missing value ini penulis menggunakan python dengan library pandas untuk melakukan Analisa atau pembersihan data jika terdapat adanya missing value pada dataset.
- Transformation
  Pada tahap transformasi data, penulis akan mengubah data kategorikal menjadi bentuk numerik untuk setiap atribut. Hal ini akan memudahkan penulis dalam memproses data dari dataset.

Tabel 2. Attribut Transformasi Tiap Kelas

Attributes	Transform
Normal	1
DoS	2
R2L	3
U2R	4
Probe	5

- Normalization
  Selama proses ini, dataset NSL KDD CUP 1999 akan mengalami transformasi atau penggabungan. Untuk mencapainya, normalisasi menggunakan Persamaan (2.1) akan diterapkan, di mana nilai atribut akan diubah skala menjadi rentang 0 hingga 1 untuk mengurangi penyebaran data. Prosedur normalisasi melibatkan penerapan persamaan normalisasi min-max, di mana nilai minimum atribut x ditandai dengan min(x), nilai maksimum dengan max(x), dan nilai hasil normalisasi dengan x'.

f. Classification
   Setelah menyelesaikan tahap transformasi, penulis sekarang akan melanjutkan ke tahap klasifikasi dataset untuk mengukur akurasi menggunakan KNN dan KNN dengan Algoritma Genetika untuk memberikan hasil analisis yang berbeda.

g. Confusion Matrix
   Confusion matrix adalah salah satu metode yang dapat digunakan untuk menghitung hasil dari metode klasifikasi, karena confusion matrix memberikan informasi yang membandingkan hasil klasifikasi yang telah dilakukan oleh sistem dengan nilai klasifikasi yang seharusnya. Penilaian hasil menggunakan confusion matrix terdiri dari beberapa tipe, yaitu:
   a) True Positive: nilai data positif yang telah diklasifikasikan dengan benar oleh sistem.

b) True Negative: nilai data negatif yang telah diklasifikasikan dengan benar oleh sistem.
c) False Positive: nilai data positif yang telah salah diklasifikasikan oleh sistem.
d) False Negative: nilai data negatif yang telah salah diklasifikasikan oleh sistem.

-. Precision : Presisi diambil berdasarkan ketiadaan informasi, dalam klasifikasi biner, presisi dapat dibuat sama dengan nilai prediksi positif, mengikuti rumus presisi.

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (1)$$

-. Recall : Recall adalah jumlah data yang berhasil diambil dari data yang sesuai dengan kueri dalam klasifikasi biner. Recall juga dikenal sebagai sensitivitas, pembentukan data diambil sesuai dengan kesepakatan pada kueri.

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

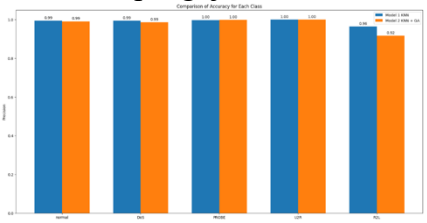
-. Accuracy : Accuracy adalah persentase dari total data yang diidentifikasi dan dihitung dengan benar.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

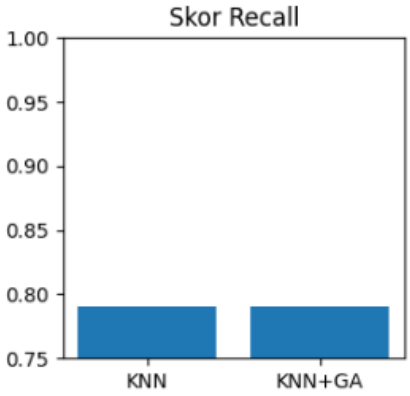
h. Accuracy
Tahap akhir adalah akurasi, di mana evaluasi ini dilakukan setelah beberapa tahap sebelumnya untuk mengambil kesimpulan tentang implementasi KNN dan KNN dengan Algoritma Genetika dari segi tingkat akurasi mereka.

HASIL DAN PEMBAHASAN PENELITIAN

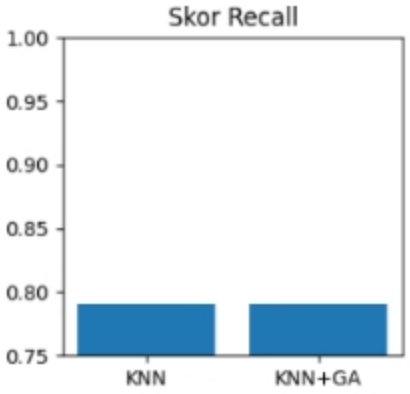
Dalam sesi ini, penulis berfokus pada evaluasi hasil Akurasi, Precission, Recall. Roc dan Confusion Matrix yang diperoleh menggunakan dua metode algoritma, yaitu KNN dan KNN dengan Algoritma Genetika. Penulis juga menunjukkan tingkat efektivitas masing-masing algoritma menggunakan Dataset KDD (Knowledge Discovery and Data Mining) dari sampel data yang dipilih secara acak untuk pelatihan dan pengujian.



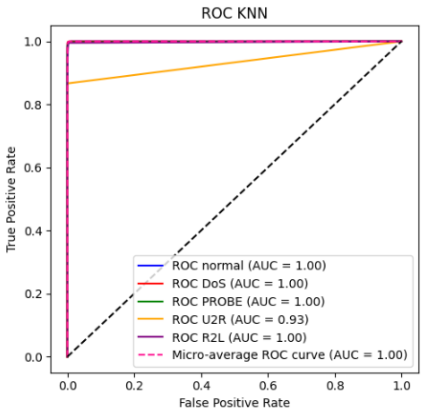
Gambar 3. Nilai akurasi dari pembandingan metode dari setiap kelas



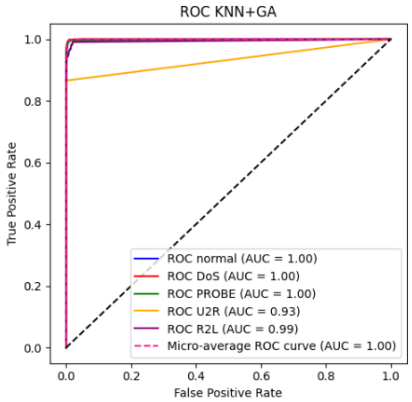
Gambar 4. Precission KNN dan KNN Genetika Algoritma



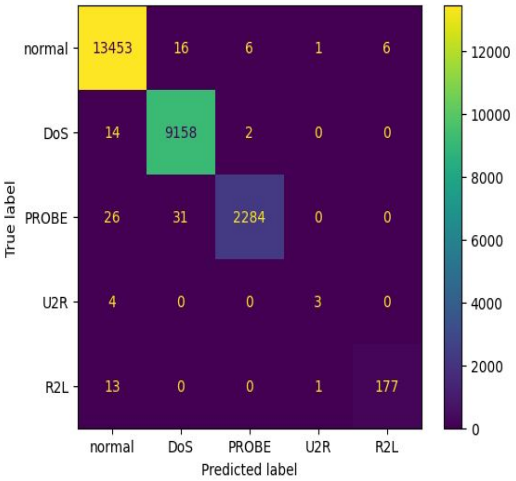
Gambar 5. Recall KNN dan KNN Genetika Algoritma



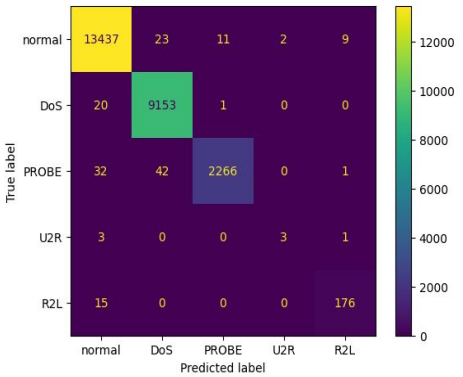
Gambar 6. ROC KNN



Gambar 7. ROC KNN dengan Genetika Algoritma



Gambar 8. Hasil Confusion Matrix dari KNN



Gambar 9. Hasil Confusion Matrix dari KNN Genetika Algoritma

PENUTUP

Kesimpulan

Kesimpulan dari hasil penggunaan metode KNN dan KNN Algoritma Genetika pada Dataset NSL KDD menunjukkan bahwa tingkat akurasi yang dicapai oleh KNN lebih tinggi dari pada KNN dengan Algoritma Genetika. Selain itu juga hasil dari seperti Precission dan Recall menunjukkan hasil yang sama satu sama lain kemudian dari hasil ROC hanya terdapat sedikit perbedaan lalu hasil dari Confusion Matrix untuk KNN juga menunjukkan nilai yang lebih baik. Semua pencapaian ini diperoleh dengan penentuan dataset, sehingga mencapai tingkat akurasi maksimum untuk KNN.

DAFTAR PUSTAKA

[1] Ardiyanti, H. (2016). Cyber-security dan tantangan pengembangannya di indonesia. Jurnal Politica Dinamika Masalah Politik Dalam Negeri dan Hubungan Internasional, 5(1).

[2] Benaddi, H., Ibrahimi, K., & Benslimane, A. (2018, October). Improving the intrusion detection system for NSL-KDD dataset based on PCA-fuzzy clustering-KNN. In 2018 6th International Conference on Wireless Networks and Mobile Communications (WINCOM) (pp. 1-6). IEEE.

[3] Dhanabal, L. &. (2015). A Study on NSL-KDD Dataset for Intrusion , 446-452.

[4] Gondohanindijo, J. (2011). Sistem Untuk Mendeteksi Adanya Penyusup (IDS: Intrusion Detection System). Majalah Ilmiah INFORMATIKA,, 2(2).

[5] Hidayanti, W. P. (2020). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada Lombok Vape On. Infotek: Jurnal Informatika dan Teknologi,, 3(2), 104-114.

[6] Houben, I. W. (1997). Genetic algorithm based k nearest neighbors, IFAC Proceedings Volumes, 30(6), 1075-1080.

[7] Imam, R. M. (2019). Deteksi Anomali Jaringan Menggunakan Hybrid Algorithm. eProceedings of Engineering, 6(2).

[8] Irsyadi, F. ((2020)). PENERAPAN MODIFIED K-NEAREST NEIGHBOR (MKNN) UNTUK KLASIFIKASI DATA SERANGAN JARINGAN KOMPUTER (NSL KDD CUP 1999). (Doctoral dissertation, Universitas Islam Negeri Sultan Syarif Kasim Riau).

[9] Li, W., Yi, P., Wu, Y., Pan, L., & Li, J. (2014). A new intrusion detection system based on KNN classification algorithm in wireless sensor network. Journal of Electrical and Computer Engineering, 2014.

[10] Liu, G., Zhao, H., Fan, F., Liu, G., Xu, Q., & Nazir, S. (2022). An enhanced intrusion detection model based on improved kNN in WSNs. Sensors, 22(4), 1407.

[11] Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 9, 381-386.

[12] Rahmad, F., Suryanto, Y., & Ramli, K. (2020, July). Performance comparison of anti-spam technology using confusion matrix classification. In IOP Conference Series: Materials Science and Engineering (Vol. 879, No. 1, p. 012076). IOP Publishing.

[13] Saravanan, A., & Bama, S. S. (2019). A review on cyber security and the fifth generation cyberattacks. Oriental journal of computer science and technology, 12(2), 50-56.

[14] Senthilnayaki, B., Venkatalakshmi, K., & Kannan, A. (2019). Intrusion detection system using fuzzy rough set feature selection and modified KNN classifier. Int. Arab J. Inf. Technol., 16(4), 746-753.

[15] Shapoorifard, H., & Shamsinejad, P. (2017). Intrusion detection using a novel hybrid method incorporating an improved KNN. Int. J. Comput. Appl, 173(1), 5-9.