Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.673-685

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31103 Link : https://dx.doi.org/10.29408/jit.v8i2.31103

Analisis Komparatif Feture Selection dan Metode Klasifikasi Intrution Detection System (IDS)

Ramli Ahmad[1*], Muhammad Saiful[2]
[1]Program Studi Teknik Komputer, Universitas Hamzanwadi
[2]Program Studi Sistem Informasi, Universitas Hamzanwadi
*ramliahmad@hamzanwadi.ac.id

Abstrak

Intrusion Detection System (IDS) adalah sistem komputer yang menganalisis lalu lintas pengguna dan data di jaringan. IDS akan memberikan sinyal peringatan kepada pengguna jaringan yang diserang penyusup. Namun, IDS masih perlu dikembangkan karena pola serangan yang terus berkembang. Pemilihan Fitur dan Klasifikasi data jaringan merupakan komponen penting dari pengembangan IDS untuk mengidentifikasi pola serangan berdasarkan data lalu lintas jaringan. Untuk mengenali pola serangan, banyak peneliti mengajukan metode Pemilihan Fitur seperti Chi-square, Gini Index, Information Gain, Relief, Gain Ratio, lan Uncertainty. Para peneliti juga telah mengusulkan metode Klasifikasi Pembelajaran Mesin seperti Random Forest, SVM, Naive Bayes, AdaBoost, dan Neural Network. Namun, di antara metode yang mereka usulkan, metode Pemilihan dan Klasifikasi Fitur, yang terbaik, belum dipelajari. Oleh karena itu, kami mengusulkan untuk melakukan studi komparatif dari metode ini untuk menemukan metode pemilihan dan klasifikasi fitur yang lebih baik. Studi ini membandingkan metode Pemilihan dan Klasifikasi Fitur untuk mengidentifikasi yang cocok untuk IDS dengan membandingkan Recall, Precision, F1, Accuracy, dan Area Under Curve (AUC) untuk setiap metode. Hasil penelitian menunjukkan bahwa metode Pemilihan Fitur Gain Rasio lebih baik daripada metode Pemilihan Fitur lainnya. Selain itu, penelitian ini menunjukkan bahwa metode klasifikasi AdaBoost lebih akurat daripada metode Klasifikasi lainnya.

Kata kunci : Chi square, Gini Index, Information Gain, Relief, Uncertainty, Gain Ratio, Random Forest, SVM, Naive Bayes, AdaBoost, Neural Network, Intrusion Detection System (IDS), Feature Selection.

Abstract

An Intrusion Detection System (IDS) is a computer system that analyzes user traffic and data on a network. IDS will provide a warning signal to network users that are attacked by intruders. However, IDS still needs to be developed due to the ever-evolving attack pattern. Feature Selection and Classification Network data are important components of IDS development to identify attack patterns based on network traffic data. To identify attack patterns, many researchers have proposed Feature Selection methods such as Chi-square, Gini Index, Information Gain, Relief, Gain Ratio, and Uncertainty. Researchers have also proposed Machine Learning Classification methods such as Random Forest, SVM, Naive Bayes, AdaBoost, and Neural Network. However, among the methods they propose, the Selection and Classification of Features method, the best, has not been studied. Therefore, we propose to conduct a comparative study of this method to find a better method of selection and classification of features. This study compares Feature Selection and Classification methods to identify those suitable for IDS by comparing Recall, Precision, F1, Accuracy, and Area Under Curve (AUC) for each method. The results showed that the Gain Ratio Feature Selection method was better than other Feature Selection methods. In addition, this study shows that the AdaBoost classification method is more accurate than other classification methods..

Keywords : *Chi square, Gini Index, Information Gain, Relief, Uncertainty, Gain Ratio, Random Forest, SVM, Naive Bayes, AdaBoost, Neural Network, Intrusion Detection System (IDS), Feature Selection.*

1. Pendahuluan

Saat ini, hampir semua aktivitas transaksi berharga, seperti perbankan dan pengiriman dokumen rahasia, dilakukan melalui jaringan internet. Namun, penyusup terus-menerus menyerang sistem koneksi jaringan Internet[1].

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.673-685

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31103 Link : https://dx.doi.org/10.29408/jit.v8i2.31103

Akibatnya, Penyusup dapat menyebabkan pencurian data atau hal-hal yang lebih buruk bagi pengguna online. Karena itu perlu ada suatu sistem yang memonitor aktivitas pengguna dan memberikan peringatan dini kepada pengguna jika ada penyusup yang mencoba masuk tanpa ijin, hal ini bisa dilakukan dengan menggunakan Sistem Deteksi Intrusi atau yang biasa dikenal dengan Intrusion Detection System yang disingkat menjadi IDS. Sistem Deteksi Intrusi (IDS) penting untuk keamanan jaringan, dan IDS digunakan untuk menganalisis dan mengontrol lalu lintas jaringan yang masuk[1]. Sebagai tambahan, metode pemilihan fitur sangat penting untuk menghapus data yang tidak relevan dan berlebihan[2].

Selanjutnya, para peneliti melakukan banyak penelitian tentang pemilihan fitur untuk IDS, yang menunjukkan bahwa pemilihan fitur yang lebih baik dapat meningkatkan akurasi suatu model[3]–[8]. Demikian pula dalam penelitian tentang metode klasifikasi, peneliti menggunakan berbagai metode untuk mengklasifikasikan serangan terhadap data IDS, tetapi metode klasifikasi mana yang lebih baik masih perlu dipelajari lebih dalam [9]–[13].

Dalam penelitian ini, kami membandingkan metode pemilihan fitur: Chi-square, Gini Index, Information Gain, Relief, Gain Ratio, dan Uncertainty untuk menemukan metode pemilihan fitur yang lebih baik daripada metode pemilihan fitur lainnya. Kami juga membandingkan metode klasifikasi seperti Random Forest, SVM, Naive Bayes, AdaBoost, dan Neural Network.

Tujuan dari penelitian ini adalah untuk menemukan metode klasifikasi yang lebih baik untuk mengenali berbagai jenis serangan di IDS..

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Dalam penelitian ini, penulis merujuk pada sejumlah penelitian terdahulu sebagai acuan. Para peneliti mengusulkan berbagai metode untuk mengatasi permasalahan IDS diantaranya:

- Selvamani menyarankan studi banding menggunakan beberapa teknik pemilihan fitur dan metode klasifikasi machine learning (ML) untuk menemukan teknik pemilihan fitur yang lebih baik di IDS [14].
- Navdeep mengusulkan teknik Support Vector Machine-CART untuk mengklasifikasikan data IDS. Para peneliti juga mengusulkan penggunaan Analisis Diskriminan Linier (LDA) untuk secara konsisten memprediksi fitur penting dalam data dimensi tinggi sehingga fitur yang tidak penting dapat dihilangkan [15].
- Saranya T mengusulkan studi banding menggunakan beberapa metode klasifikasi data IDS menggunakan algoritma ML dengan penambahan Linear Discriminant Analysis (LDA), Classification and Regression Trees (CART), dan Random Forest [16].

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.673-685

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31103 Link : https://dx.doi.org/10.29408/jit.v8i2.31103

- Kushwaha menyarankan metode statistik untuk mengklasifikasikan data IDS untuk menemukan algoritma yang paling relevan untuk pemilihan fitur dari antara 41 vektor koneksi atribut [17].

2.2. Landasan Teori

Beberapa teori yang dibutuhkan dalam penelitian dijabarkan secara terperinci sebagai berikut :

1. Deskripsi Dataset KDD Cup 99

Pada tahun 1999, Laboratorium MIT Lincoln diumumkan sebagai Kumpulan Data KDD Cup 99. Tuple dataset berisi 41 variabel karakteristik dan satu label yang menunjukkan apakah 'normal' atau 'serangan' dengan total 494.021 catatan sebagai 10% set data pelatihan. Empat kategori jenis serangan dalam dataset KDD Cup 99 sebagai berikut:

- Denial of Service Attack (DOS): Penyerang mencoba membuat sumber daya tidak dapat diakses oleh pengguna resminya dalam serangan semacam ini [18].
- User to Root Attack (U2R): Di antara berbagai jenis serangan, serangan siber adalah yang paling ditakuti [19].
- Remote to Local Attack (R2L): Mendapatkan akses lokal ke komputer korban adalah tujuan utama dari jenis serangan ini [20].
- Probing Attack (PROBE): Penyerang mengintip jaringan target untuk mendapatkan sistem langsung, port terbuka, dan sebagainya[21].

Nominal serangan individu digunakan untuk membagi kumpulan data pelatihan dan pengujian. Dataset KDD Cup 99 dibagi menjadi 5 kategori serangan yaitu Probe, U2R, DOS, R2L, dan Normal. Kelima kategori serangan tersebut terdiri dari 23 serangan distribusi [22].

Tabel 1. Kategori Serangan KDD Cup 99

Kategori	Nama Serangan
Normal	Normal
Probe	Satan, portsweep, Nmap, ipsweep
DoS	Teardrop, Smurf, Pod, Neptune, land, back.
U2R	Perl, rootkit, loadmodule, buffer_overflow
R2L	Warezmaster, warezclient, spy, imap, phf, multihop, guess_passwd, ftp-write

Berdasarkan Tabel 1, nama serangan seperti normal, teardrop, portsweep, smurf, satan, Neptunus, back, warezclient, ipsweep, memiliki jumlah data yang signifikan untuk diproses. Sementara itu, nama serangan seperti nmap, pod, buffer_overflow, dan guess_passwd, memiliki jumlah data yang tidak terlalu baik tetapi masih dapat diproses.

2. Metode Feature Selection

Metode pemilihan fitur yang digunakan dalam penelitian ini antara lain :

- Chi Square adalah teknik analisis statistik untuk menentukan tingkat signifikansi statistik dari aturan asosiasi serta menentukan perbedaan antara distribusi yang diukur dan diprediksi [23].

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.673-685

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31103 Link : https://dx.doi.org/10.29408/jit.v8i2.31103

- Relief adalah pendekatan klasifikasi biner dengan kualitas diskrit atau numerik yang dibuat untuk tujuan memecahkan masalah klasifikasi biner [6].
- Information Gain merupakan konsep perolehan informasi mengacu pada seberapa banyak informasi yang dapat dikontribusikan oleh suatu fitur pada sistem kategorisasi [24].
- Uncertainty merupkan sutau IDS di bidang kinerja, ketidakpastian sangat penting dalam hal nilai karakteristik seperti permintaan sumber daya layanan dan waktu servis perangkat keras [25].
- Gain Ratio adalah variasi pengurangan bias dari perolehan informasi. Saat memilih atribut, ini mengevaluasi jumlah dan ukuran cabang [26].
- Gini Index adalah alat pemilihan fitur populer yang telah menerima banyak perhatian akademis. Ini memiliki kompleksitas komputasi yang rendah dan dapat meningkatkan kinerja klasifikasi [4].

Dalam penelitian kami, kami membandingkan dan menentukan pemilihan fitur mana yang terbaik menggunakan semua metode ini.

3. Metode Klassifikasi

Metode klasifikasi Machine Learning dalam penelitian ini meliputi:

- AdaBoost classifier: AdaBoost adalah metode untuk konstruksi ansambel berulang. Ini membangun pengklasifikasi yang kuat dengan menggabungkan banyak pengklasifikasi sub optimal [27].
- Random Forests Classifier: Model Random Forest dibangun dengan menggabungkan setiap pohon yang baik menjadi satu model [28].
- Neural Network Classifier: Jaringan saraf adalah kumpulan neuron yang terhubung dalam jaringan atau sirkuit, atau dalam arti yang lebih modern, jaringan saraf buatan [29].
- Naive Bayes classifier: Naive Bayes adalah pendekatan prediksi probabilistik langsung berdasarkan penerapan teorema atau aturan [30].
- SVM classifier: SVM, atau Support Vector Machine adalah pendekatan yang sering digunakan untuk masalah Regresi dan Klasifikasi yang membutuhkan Supervised Learning [31].

4. Parameter Untuk Evaluasi

Matriks evaluasi yang digunakan dalam penelitian ini meliputi:

- Recall: adalah perbandingan antara True Positive (TP) dengan jumlah data yang sebenarnya positif. Dan itu bisa ditulis secara matematis:

$$\text{Recall} = \frac{(TP)}{(TP+FN)} \quad (1)$$

- Accuracy: adalah kedekatan derajat antara nilai prediksi dan aktual. Dan itu bisa ditulis secara matematis:

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.673-685

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31103 Link : https://dx.doi.org/10.29408/jit.v8i2.31103

$$\text{Acc} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (2)$$

- Precision: adalah perbandingan antara True Positive (TP) dengan jumlah data yang diprediksi positif. Atau bisa juga ditulis secara matematis:

$$\text{Precision} = \frac{(TP)}{(TP+FP)} \quad (3)$$

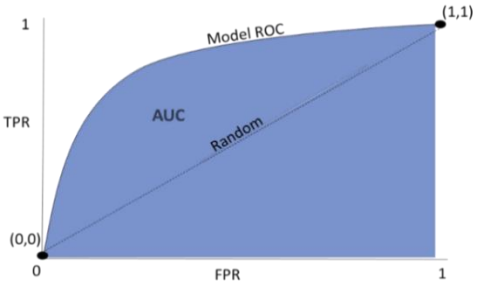
- F1 Score: adalah makna harmonik dari presisi dan ingatan. Yang bisa ditulis secara matematis:

$$\text{F1} = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (4)$$

- AUC: adalah makna harmonik dari presisi dan ingatan. Yang bisa ditulis secara matematis.

$$\text{TPR} = \frac{(TP)}{(TP+FN)} \quad \& \quad \text{FPR} = \frac{(FP)}{(FP+TN)} \quad (5)$$

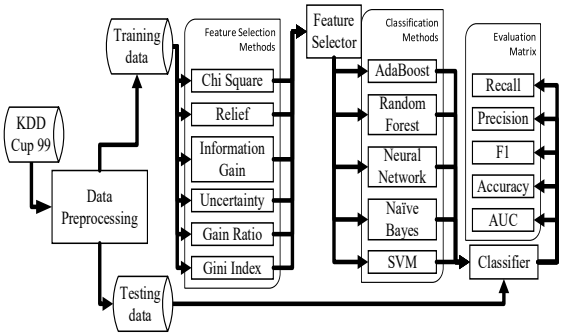
Area di bawah kurva ROC. Ini mengukur area 2D kurva ROC penuh di bawahnya. AUC memperhitungkan keseluruhan klasifikasi yang tersedia.



Gambar 1. AUC curve [32]

3. Metode Penelitian

Dalam penelitian kami, kami mengusulkan metode menggunakan dataset KDD Cup 99, dan fase pra-pemrosesan data adalah membagi data menjadi dua, yaitu: data pelatihan dan pengujian. Kemudian, data pelatihan diolah menggunakan 6 metode pemilihan fitur, yaitu: Chi Square, Relief, Information Gain, Uncertainty, Gain Ratio, dan Gini Index, yang dikombinasikan dengan data pengujian. Selanjutnya, hasil yang diperoleh dalam metode pemilihan fitur diolah menggunakan lima metode klasifikasi atau pengklasifikasi, yaitu: AdaBoost, Random Forest, Neural Network, Naïve Bayes, dan SVM. Pada fase akhir kami mengevaluasi pengklasifikasi menggunakan penarikan, presisi, skor F1, akurasi, dan AUC. Untuk detail selengkapnya, lihat Gambar 2.



Gambar 2. Kerangka Experimental

Setelah proses RapidMiner selesai, hasilnya adalah sebagai berikut: chi square 24 Fitur, relief 17 Fitur, information gain 27 Fitur, uncertainty 28 Fitur, gain ratio 34 Fitur, and Gini Index 17 Fitur. Untuk lebih jelasnya, lihat Tabel 2.

Tabel 2. Fitur yang telah dipilih berdasarkan metode Feature Selection

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.673-685

e-ISSN 2614-8773

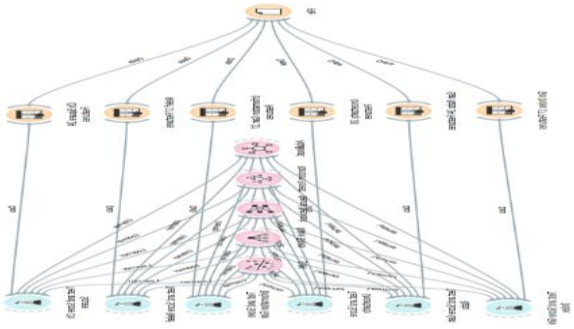
DOI : 10.29408/jit.v8i2.31103 Link : https://dx.doi.org/10.29408/jit.v8i2.31103

No	Fitur KDD Cup 99	Chi Square	Relief	Information Gain	Uncertainty	Gain Ratio	Gini Index
1	lnum_shells	1		1	1	1	
2	lnum_file_creations	1		1	1	1	
3	service	1	1	1	1	1	
4	lnum_compromised	1		1	1	1	
5	Flag	1		1	1	1	
6	count	1	1	1	1	1	1
7	srv_count	1	1	1	1	1	1
8	protocol_type	1	1	1	1	1	1
9	dst_host_same_src_port_rate	1	1	1	1	1	1
10	logged_in	1	1	1	1	1	1
11	dst_host_srv_diff_host_rate	1	1	1	1	1	1
12	diff_srv_rate	1	1	1	1	1	
13	serror_rate	1		1	1	1	
14	dst_host_srv_serror_rate	1		1	1	1	1
15	dst_host_serror_rate	1		1	1	1	1
16	srv_serror_rate	1		1	1	1	1
17	same_srv_rate	1	1	1	1	1	
18	lnum_root	1			1	1	
19	lroot_shell	1		1	1	1	
20	dst_host_count	1	1	1	1	1	1
21	srv_diff_host_rate	1	1	1	1	1	1
22	dst_host_srv_count	1	1	1	1	1	1
23	dst_host_same_srv_rate	1	1	1	1	1	1
24	dst_bytes	1	1	1	1	1	1
25	src_bytes		1	1	1	1	1
26	dst_host_diff_srv_rate		1	1	1	1	1
27	hot			1	1	1	
28	is_guest_login				1	1	
29	lnum_access_files		1			1	
30	dst_host_srv_rerror_rate					1	
31	rerror_rate					1	
32	srv_rerror_rate					1	
33	dst_host_rerror_rate					1	
34	duration			1		1	1
35	land						
36	wrong_fragment						
37	urgent						
38	num_failed_logins						
39	lsu_attempted						
40	lnum_outbound_cmds						
41	is_host_login						

Fase kedua kami menggunakan Orange untuk menemukan dan menentukan pemilihan fitur terbaik dan metode klasifikasi dengan membandingkan 6 pemilihan fitur dan 5 metode klasifikasi. Untuk detail lebih lanjut dapat dilihat pada Gambar 3.

Proses penentuan metode klasifikasi terbaik ditunjukkan pada Gambar 3. Pada tahap kedua ini, fitur-fitur yang dipilih oleh masing-masing dari enam metode pemilihan fitur digunakan sebagai input untuk lima metode klasifikasi, sehingga metode klasifikasi menerima 6 data input yang berbeda, yaitu: chi square 24 fitur, relief 17 fitur, gain informasi 27 fitur, ketidakpastian 28 fitur, gain ratio 34 fitur, Gini Index 17 fitur. Untuk lebih jelasnya lihat Gambar 3.

Untuk mnedapatkan nilai AUC, Akurasi, F1, Presisi, dan Recall, lima metode klasifikasi diantaranya AdaBoost, Random Forest, Neural Network, Gain Ratio, dan Gini Index diproses melalui evaluasi metrik menggunakan pengujian data. penjelasan seluruh proses dapat dilihat pada Gambar 3.



Gambar 3. Alur proses untuk menghitung pengklasifikasi terbaik dan pemilihan fitur

4. Hasil dan Pembahasan

Bagian berikut menjelaskan eksperimen kami dan hasil yang kami peroleh. Selain itu, kami

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.673-685

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31103 Link : https://dx.doi.org/10.29408/jit.v8i2.31103

menggunakan alat RapidMiner dan Orange untuk menggambarkan hasil kami.

4.1. Hasil Penelitian

Pengamatan ini akan menjelaskan secara rinci hasil Recall, Precision, F1, Accuracy, dan AUC pada setiap model Feature Selection, serta menganalisis nilai yang diperoleh untuk menentukan model Feature Selection mana yang memiliki performa terbaik dan classifier mana yang memiliki performa terbaik. Dalam penelitian kami, kami menetapkan batas nilai terendah 0,75 dan batas nilai tertinggi 1,00, yang kami gunakan sebagai referensi untuk metode pengklasifikasi dan nilai metode pemilihan fitur untuk menemukan metode pengklasifikasi terbaik dan metode pemilihan fitur.

Tabel 3. Hasil Pemilihan FItur dan Metode Klasifikasi

Metode Klassifikasi	Evaluation Matrix	Metode Feature Selection					
		Chi Square	Relief	Information Gain	Uncertainty	Gain Ratio	Gini Index
AdaBoost	Recall	0.73	0.70	0.73	0.74	0.75	0.69
	Precision	0.77	0.74	0.77	0.77	0.79	0.73
	f1-score	0.73	0.71	0.74	0.74	0.70	0.69
	Accuracy	1.00	0.99	1.00	1.00	0.99	0.99
	AUC	0.95	0.94	0.94	0.95	0.94	0.94
Random Forest	Recall	0.69	0.67	0.69	0.70	0.71	0.66
	Precision	0.76	0.73	0.76	0.76	0.78	0.72
	f1-score	0.70	0.67	0.70	0.70	0.66	0.66
	Accuracy	0.99	0.99	0.99	0.99	0.99	0.98
	AUC	0.94	0.93	0.93	0.94	0.93	0.93
Neural Network	Recall	0.71	0.68	0.71	0.72	0.73	0.67
	Precision	0.74	0.71	0.74	0.74	0.76	0.70
	f1-score	0.71	0.68	0.71	0.71	0.67	0.67
	Accuracy	1.00	0.99	1.00	1.00	0.99	0.99
	AUC	0.95	0.95	0.95	0.95	0.95	0.95
Naïve Bayes	Recall	0.70	0.68	0.70	0.71	0.72	0.67
	Precision	0.59	0.57	0.60	0.60	0.62	0.56
	f1-score	0.60	0.57	0.60	0.60	0.56	0.56
	Accuracy	0.97	0.97	0.97	0.97	0.97	0.96
	AUC	0.95	0.95	0.95	0.95	0.95	0.95
SVM	Recall	0.66	0.63	0.66	0.67	0.68	0.62
	Precision	0.68	0.66	0.69	0.69	0.71	0.65
	f1-score	0.65	0.62	0.65	0.65	0.61	0.61
	Accuracy	0.98	0.98	0.98	0.98	0.98	0.97
	AUC	0.95	0.95	0.95	0.95	0.95	0.95

1. Hasil Metode Feature Selection

Menurut Tabel 3. Hasil dari metode pemilihan fitur dapat dilihat sebagai berikut:

a. Hasil Chi Square

Dalam metode pemilihan fitur "Chi Square", tidak ada nilai "recall" yang memiliki nilai ≥ 0,75. Pada "Precision", terdapat dua metode klasifikasi yang memiliki nilai 0,75, yaitu "AdaBoost" dengan nilai 0,77 dan "Random Forest" dengan nilai 0,76. Dalam "skor f1", tidak ada metode klasifikasi yang memiliki nilai ≥ 0,75. Dalam "Akurasi", semua metode klasifikasi memiliki nilai 0,75, yaitu: "AdaBoost" dengan nilai 1,00, "Random Forest" dengan nilai 0,99, "Neural Network" dengan nilai 1,00, "Nave Bayes" dengan nilai 0,97, dan "SVM" dengan nilai 0,98. Dalam "AUC", semua metode klasifikasi memiliki nilai 0,75, yaitu: "AdaBoost" dengan nilai 0,95, "Random Forest" dengan nilai 0,94, "Neural Network" dengan nilai 0,95, "Nave Bayes" dengan nilai 0,95, dan "SVM" dengan nilai 0,95.

b. Hasil Relief

Dalam metode pemilihan fitur "Relief", tidak ada nilai "recall" yang memiliki nilai ≥ 0,75. Dalam

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.673-685

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31103 Link : https://dx.doi.org/10.29408/jit.v8i2.31103

"Presisi", tidak ada metode klasifikasi yang memiliki nilai ≥ 0,75. Dalam "skor f1", juga tidak ada metode klasifikasi tunggal yang memiliki nilai ≥ 0,75. Dalam "Akurasi", semua metode klasifikasi memiliki nilai 0,75, yaitu: "AdaBoost" dengan nilai 0,99, "Random Forest" dengan nilai 0,99, "Neural Network" dengan nilai 0,99, "Nave Bayes" dengan nilai 0,97, dan "SVM" dengan nilai 0,98. Dalam "AUC", semua metode klasifikasi memiliki nilai ≥ 0,75, yaitu: "AdaBoost" dengan nilai 0,95, "Random Forest" dengan nilai 0,94, "Neural Network" dengan nilai 0,95, "Nave Bayes" dengan nilai 0,97, dan "SVM" dengan nilai 0,98. Ini menunjukkan bahwa metode klasifikasi "AdaBoost" berkinerja lebih baik daripada yang lain dalam hal nilai rata-rata.

c. Hasil Information Gain

Dalam metode pemilihan fitur "Keuntungan Informasi", tidak ada nilai "penarikan" yang memiliki nilai ≥ 0,75. Dalam "Precision", terdapat dua metode klasifikasi yang memiliki nilai ≥ 0,75, yaitu "AdaBoost" dengan nilai 0,77 dan "Random Forest" dengan nilai 0,76. Dalam "skor f1", tidak ada metode klasifikasi yang memiliki nilai ≥ 0,75. Dalam "Akurasi", semua metode klasifikasi memiliki nilai ≥ 0,75, yaitu: "AdaBoost" dengan nilai 1,00, "Random Forest" dengan nilai 0,99, "Neural Network" dengan nilai 1,00, "Nave Bayes" dengan nilai 0,97, dan "SVM" dengan nilai 0,98. Dalam "AUC", semua metode klasifikasi memiliki nilai ≥ 0,75, yaitu: "AdaBoost" dengan nilai 0,94, "Random Forest" dengan nilai 0,93, "Neural Network" dengan nilai 0,95, "Nave Bayes" dengan nilai 0,95, dan "SVM" dengan nilai 0,95. Ini menunjukkan bahwa metode klasifikasi "AdaBoost" berkinerja lebih baik daripada yang lain dalam hal nilai rata-rata.

d. Hasil Uncertainty

Dalam metode pemilihan fitur "ketidakpastian", tidak ada nilai "recall" yang memiliki nilai ≥ 0,75. Pada "Precision", terdapat dua metode klasifikasi yang memiliki nilai ≥ 0,75, yaitu "AdaBoost" dengan nilai 0,77 dan "Random Forest" dengan nilai 0,76. Dalam "skor f1", tidak ada metode klasifikasi yang memiliki nilai ≥ 0,75. Dalam "Akurasi", semua metode klasifikasi memiliki nilai ≥ 0,75, yaitu: "AdaBoost" dengan nilai 1,00, "Random Forest" dengan nilai 0,99, "Neural Network" dengan nilai 1,00, "Naïve Bayes" dengan nilai 0,97, dan "SVM" dengan nilai 0,98. Dalam "AUC", semua metode klasifikasi memiliki nilai ≥ 0,75, yaitu: "AdaBoost" dengan nilai 0,95, "Random Forest" dengan nilai 0,94, "Neural Network" dengan nilai 0,95, "Nave Bayes" dengan nilai 0,95, dan "SVM" dengan nilai 0,95. Ini menunjukkan bahwa metode klasifikasi "AdaBoost" berkinerja lebih baik daripada yang lain dalam hal nilai rata-rata.

e. Hasil Gain Ratio

Dalam metode pemilihan fitur "Gain Ratio", hanya metode klasifikasi "AdaBoost" yang memiliki nilai "Recall", yaitu nilai ≥ 0,75. Dalam "Precision",

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.673-685

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31103 Link : https://dx.doi.org/10.29408/jit.v8i2.31103

terdapat tiga metode klasifikasi yang memiliki nilai ≥ 0,75, yaitu: "AdaBoost" dengan nilai 0,79, "Random Forest" dengan nilai 0,78, dan "Neural Network" dengan nilai 0,76. Pada "skor f1", tidak ada metode klasifikasi yang memiliki nilai ≥ 0,75. Dalam "Akurasi", semua metode klasifikasi memiliki nilai ≥ 0,75, yaitu: "AdaBoost" dengan nilai 0,99, "Random Forest" dengan nilai 0,99, "Neural Network" dengan nilai 0,99, "Naive Bayes" dengan nilai 0,97, dan "SVM" dengan nilai 0,98. Dalam "AUC", semua metode klasifikasi memiliki nilai ≥ 0,75, yaitu: "AdaBoost" dengan nilai 0,94, "Random Forest" dengan nilai 0,93, "Neural Network" dengan nilai 0,95, "Naïve Bayes" dengan nilai 0,95, dan "SVM" dengan nilai 0,95. Ini menunjukkan bahwa metode klasifikasi "AdaBoost" berkinerja lebih baik daripada yang lain dalam hal nilai rata-rata.

f. Hasil Gini Index

Ing metode pemilihan fitur "Gini Index", ora ana nilai "recall" sing duwe nilai ≥ 0.75. Dalam "Presisi", tidak ada metode klasifikasi yang memiliki nilai ≥ 0,75. Dalam "skor f1", juga tidak ada metode klasifikasi tunggal yang memiliki nilai ≥ 0,75. Dalam "Akurasi", semua metode klasifikasi memiliki nilai ≥ 0,75, yaitu: "AdaBoost" dengan nilai 0,99, "Random Forest" dengan nilai 0,98, "Neural Network" dengan nilai 0,99, "Naïve Bayes" dengan nilai 0,96, dan "SVM" dengan nilai 0,97. Dalam "AUC", semua metode klasifikasi memiliki nilai ≥ 0,75, yaitu: "AdaBoost" dengan nilai 0,94, "Random Forest" dengan nilai 0,93, "Neural Network" dengan nilai 0,95, "Naïve Bayes" dengan nilai 0,95, dan "SVM" dengan nilai 0,95. Ini menunjukkan bahwa metode klasifikasi "AdaBoost" berkinerja lebih baik daripada yang lain dalam hal nilai rata-rata.

2. Hasil Metode Klassifikasi

Berdasarkan Tabel 3. Dapat dilihat bahwa hasil percobaan yang diperoleh pada metode klasifikasi adalah sebagai beikut:

a. Hasil AdaBoost

Dalam metode klasifikasi "AdaBoost", hanya metode pemilihan fitur "Rasio Penguatan" yang memiliki nilai "Recall" dengan nilai ≥ 0,75. Dalam "Presisi", terdapat empat metode pemilihan fitur yang memiliki nilai ≥ 0,75, yaitu: "Chi Square" dengan nilai 0,77, "Information Gain" dengan nilai 0,77, "Uncertainty" dengan nilai 0,77, dan "Gain Ratio" dengan nilai 0,79. Dalam "F1-Score", tidak ada metode pemilihan fitur yang memiliki nilai ≥ 0,75. Dalam "Akurasi", semua metode pemilihan fitur memiliki nilai ≥ 0,75, yaitu: "Chi Square" dengan nilai 1,00, "Relief" dengan nilai 0,99, "Information Gain" dengan nilai 1,00, "Uncertainty" dengan nilai 1,00, "Gain Ratio" dengan nilai 0,99, dan "Gini Index" dengan nilai 0,99. Dalam "AUC", semua metode pemilihan fitur memiliki nilai ≥ 0,75, yaitu: "Chi Square" dengan nilai 0,95, "Relief" dengan nilai 0,94, "Information Gain" dengan nilai 0,94, "Uncertainty" dengan nilai 0,95, "Gain Ratio" dengan nilai 0,94, dan

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.673-685

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31103 Link : https://dx.doi.org/10.29408/jit.v8i2.31103

"Gini Index" dengan nilai 0,94. Ini menunjukkan bahwa nilai rata-rata "rasio penguatan" lebih besar dari nilai rata-rata metode pemilihan fitur lainnya.

b. Hasil Random Forest

Dalam metode klasifikasi "Random Forest", tidak ada metode pemilihan fitur yang memiliki nilai "Recall" dengan nilai ≥ 0,75. Ada empat metode seleksi fitur yang memiliki nilai ≥ 0,75, yaitu: "Chi Square" dengan nilai 0,76, "Information Gain" dengan nilai 0,76, "Uncertainty" dengan nilai 0,76, dan "Gain Ratio" dengan nilai 0,78. pada skor f1 Tidak ada metode pemilihan fitur yang memiliki nilai ≥ 0,75. Pada "Akurasi", semua pilihan fitur memiliki nilai ≥ 0,75, yaitu: "Chi Square" dengan nilai 0,99, "Relief" dengan nilai 0,99, "Information Gain" dengan nilai 0,99, "Uncertainty" dengan nilai 0,99, "Gain Ratio" dengan nilai 0,99, dan "Gini Index" dengan nilai 0,98. Ing "AUC", kabeh metode pemilihan fitur duwe nilai ≥ 0,75, yaiku: "Chi Square" kanthi nilai 0,94, "Relief" kanthi nilai 0,93, "Information Gain" kanthi nilai 0,93, "Uncertainty" kanthi nilai 0,94, "Gain Ratio" kanthi nilai 0,93, lan "Gini Index" kanthi nilai 0,93. Ini menunjukkan bahwa nilai rata-rata "rasio penguatan" lebih besar dari nilai rata-rata metode pemilihan fitur lainnya.

c. Hasil Neural Network

Dalam metode klasifikasi "Neural Network", tidak ada metode pemilihan fitur yang memiliki nilai "Recall" dengan nilai ≥ 0,75. Hanya metode pemilihan fitur "Rasio Penguatan" yang memiliki nilai "Presisi" ≥ 0,75 dengan nilai 0,76. Dalam "F1-Score", tidak ada metode pemilihan fitur yang memiliki nilai ≥ 0,75. Dalam "Akurasi", semua metode pemilihan fitur memiliki nilai ≥ 0,75, yaitu: "Chi Square" dengan nilai 1,00, "Relief" dengan nilai 0,99, "Information Gain" dengan nilai 1,00, "Uncertainty" dengan nilai 1,00, "Gain Ratio" dengan nilai 0,99, dan "Gini Index" dengan nilai 0,99. Dalam "AUC", semua metode pemilihan fitur memiliki nilai ≥ 0,75, yaitu: "Chi Square" dengan nilai 0,95, "Relief" dengan nilai 0,95, "Information Gain" dengan nilai 0,95, "Uncertainty" dengan nilai 0,95, "Gain Ratio" dengan nilai 0,95, dan "Gini Index" dengan nilai 0,95. Ini menunjukkan bahwa nilai rata-rata "rasio penguatan" lebih besar dari nilai rata-rata metode pemilihan fitur lainnya.

d. Hasil Naïve Bayes

Dalam metode klasifikasi "Naïve Bayes", tidak ada metode pemilihan fitur yang memiliki nilai "Recall" dengan nilai ≥ 0,75. Demikian juga, dalam "presisi", tidak ada metode pemilihan fitur yang memiliki nilai ≥ 0,75. Demikian juga, pada "skor f1", tidak ada metode pemilihan fitur yang memiliki nilai ≥ 0,75. Dalam "Akurasi", semua metode pemilihan fitur memiliki nilai ≥ 0,75, yaitu: "Chi Square" dengan nilai 0,97, "Relief" dengan nilai 0,97, "Information Gain" dengan nilai 0,97, "Uncertainty" dengan nilai 0,97, "Gain Ratio" dengan nilai 0,97, dan "Gini Index" dengan nilai

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.673-685

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31103 Link : https://dx.doi.org/10.29408/jit.v8i2.31103

0,96. Dalam "AUC", semua metode pemilihan fitur memiliki nilai ≥ 0,75, yaitu: "Chi Square" dengan nilai 0,95, "Relief" dengan nilai 0,95, "Information Gain" dengan nilai 0,95, "Uncertainty" dengan nilai 0,95, "Gain Ratio" dengan nilai 0,95, dan "Gini Index" dengan nilai 0,95. Ini menunjukkan bahwa nilai rata-rata "rasio penguatan" lebih besar dari nilai rata-rata metode pemilihan fitur lainnya.

e. Hasil SVM

Dalam metode klasifikasi "SVM", tidak ada metode pemilihan fitur yang memiliki nilai "Recall" dengan nilai ≥ 0,75. Demikian juga, dalam "presisi", tidak ada metode pemilihan fitur yang memiliki nilai ≥ 0,75. Demikian juga, pada "skor f1", tidak ada metode pemilihan fitur yang memiliki nilai ≥ 0,75. Dalam "Akurasi", semua metode pemilihan fitur memiliki nilai ≥ 0,75, yaitu: "Chi Square" dengan nilai 0,98, "Relief" dengan nilai 0,98, "Information Gain" dengan nilai 0,98, "Uncertainty" dengan nilai 0,98, "Gain Ratio" dengan nilai 0,98, dan "Gini Index" dengan nilai 0,97. Dalam "AUC", semua metode pemilihan fitur memiliki nilai ≥ 0,75, yaitu: "Chi Square" dengan nilai 0,95, "Relief" dengan nilai 0,95, "Information Gain" dengan nilai 0,95, "Uncertainty" dengan nilai 0,95, "Gain Ratio" dengan nilai 0,95, dan "Gini Index" dengan nilai 0,95. Ini menunjukkan bahwa nilai rata-rata "rasio penguatan" lebih besar dari nilai rata-rata metode pemilihan fitur lainnya.

4.2. Pembahasan

Berdasarkan "Metode Pemilihan Fitur" dan "Metode Klasifikasi atau Pengklasifikasi yang diperkenalkan" di bab sebelumnya, kami menerapkan proses dua fase: fase pertama kami menggunakan RapidMiner untuk menemukan dan menentukan metode Pemilihan Fitur atau pemilih fitur terbaik. Kemudian pada fase kedua kita menggunakan Orange untuk menemukan dan menentukan metode klasifikasi atau pengklasifikasi terbaik. Data Pelatihan KDD Cup 99 digunakan sebagai input metode seleksi enam fitur, kemudian diperoleh hasil masing-masing metode seleksi enam fitur, dengan fitur tersebut dipilih dengan pemeringkatan masing-masing metode seleksi enam fitur

5. Kesimpulan

Kesimpulannya, kami berhasil mengkategorikan jenis serangan terhadap IDS menjadi lima bagian, yaitu Normal, Probe, DoS, U2R, dan R2L. Kami mengkategorikan lima jenis serangan ini dari 23 jenis serangan berdasarkan kumpulan data asli.

Analisis Komparatif berdasarkan Pemilihan dan Klasifikasi Fitur dilakukan untuk mengidentifikasi pengklasifikasi terbaik di antara 5 pengklasifikasi (AdaBoost, Random Forest, Neural Network, Naive bayes, dan SVM). Gain Ratio dianggap sebagai metode pemilihan fitur paling Bagus pada recall, presisi, F1 Score, sedangkan Indeks Gini kurang mengaruhi metode pemilihan fitur.

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.673-685

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31103 Link : https://dx.doi.org/10.29408/jit.v8i2.31103

Berdasarkan Akurasi, Chi Square, Gain Informasi, dan uncertainty berdasarkan metode pemilihan fitur labih bagus, sedngakan Gini Index merupakan pemilih fitur paling rendah.

Penelitian lebih lanjut tentang efektivitas komputasi model pembelajaran mesin dan pemilihan fitur dapat dilakukan.

6. Daftar Pustaka

[1] Z. K. Ibrahim, M. Y. Thanon, Z. Khalid, and M. Thanoun, Performance Comparison of Intrusion Detection System Using Three Different Machine Learning Algorithms. 2021.

[2] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, “A Deep Learning Approach to Network Intrusion Detection,” IEEE Trans. Emerg. Top. Comput. Intell., vol. 2, no. 1, pp. 41–50, 2018, doi: 10.1109/TETCI.2017.2772792.

[3] I. Sumaiya Thaseen and C. Aswani Kumar, “Intrusion Detection Model using Fusion of Chi-square Feature Selection and Multi Class SVM,” J. King Saud Univ. - Comput. Inf. Sci., vol. 29, no. 4, pp. 462–472, 2017, doi: https://doi.org/10.1016/j.jksuci.2015.12.004.

[4] L. Wu and Y. Wang, “Fusing Gini Index and Term Frequency for Text Feature Selection,” IEEE Third Int. Conf. Multimed. Big Data, 2017.

[5] M. T.-S. Noelia S´anchez-Maro˜no, Amparo Alonso-Betanzos, “Filter Methods for Feature Selection. A Comparative Study,” 8th Int. Conf. Intell. Data Eng. Autom. Learn. - IDEAL, Birmingham, UK, vol. 4881, pp. 178–187, 2007, doi: https://doi.org/10.1007/978-3-540-77226-2_19.

[6] R. Fu, P. Wang, Y. Gao, and X. Hua, “A new feature selection method based on relief and SVM-RFE,” Int. Conf. Signal Process. Proceedings, ICSP, 2014.

[7] A. Binbusayyis and T. Vaiyapuri, “Identifying and Benchmarking Key Features for Cyber Intrusion Detection: An Ensemble Approach,” IEEE Access, vol. PP, pp. 106495–106513, Jul. 2019, doi: 10.1109/ACCESS.2019.2929487.

[8] C. Callegari, S. Giordano, and M. Pagano, “An information-theoretic method for the detection of anomalies in network traffic,” Comput. Secur., vol. 70, pp. 351–365, 2017, doi: 10.1016/j.cose.2017.07.004.

[9] K. Hassine, A. Erbad, and R. Hamila, “Important complexity reduction of random forest in multi-classification problem,” 2019 15th Int. Wirel. Commun. Mob. Comput. Conf. IWCMC 2019, vol. 14, 2019.

[10] A. N. Jaber and S. U. Rehman, “FCM–SVM based intrusion detection system for cloud computing environment,” Cluster Comput., vol. 23, no. 4, pp. 3221–3231, 2020, doi: 10.1007/s10586-020-03082-6.

[11] M. K. Hooshmand, “Machine Learning Based Network Anomaly Detection,” Int. J. Recent Technol. Eng., vol. 8, no. 4, pp. 542–548, 2019, doi: 10.35940/ijrte.d7271.118419.

[12] A. I. Madbouly and T. M. Barakat, “Enhanced Relevant Feature Selection Model for Intrusion Detection Systems,” Int. J. Intell. Eng. Informatics, vol. 4, p. 21, 2016, doi: 10.1504/IJIEI.2016.074499.

[13] H. H. He, X. Sun, H. H. He, G. Zhao, L. He, and J. Ren, “A Novel Multimodal-Sequential Approach Based on Multi-View Features for Network Intrusion Detection,” IEEE Access, vol. 7, pp. 183207–183221, 2019, doi: 10.1109/ACCESS.2019.2959131.

[14] D. Selvamani and V. Selvi, “A Comparative Study on the Feature Selection Techniques for Intrusion Detection System,” Asian J. Comput. Sci. Technol., vol. 8, pp. 42–47, Feb. 2019, doi: 10.51983/ajcst-2019.8.1.2120.

[15] I. S. V. Navdeep, “Linear Discriminant Analysis based Hybrid SVM- CART for Intrusion Detection System,” Int. J. Eng. Dev. Res., vol. 3, no. 4, 2015.

Infotek : Jurnal Informatika dan Teknologi
Vol. 8 No. 2, Juli 2025
Hal.673-685

e-ISSN 2614-8773

DOI : 10.29408/jit.v8i2.31103 Link : https://dx.doi.org/10.29408/jit.v8i2.31103

[16] T. Saranya, S. Sridevi, C. Deisy, T. D. Chung, and M. K. A. A. Khan, "Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review," Procedia Comput. Sci., 2020, doi: https://doi.org/10.1016/j.procs.2020.04.133.

[17] P. Kushwaha, H. Buckchash, and B. Raman, "Anomaly based Intrusion Detection using Filter based Feature Selection on KDD-CUP 99," Int. Conf. Proceedings/TENCON, pp. 839–844, 2017.

[18] M. Ahmed, A. Naser Mahmood, and J. Hu, "A survey of network anomaly detection techniques," J. Netw. Comput. Appl., vol. 60, pp. 19–31, 2016, doi: 10.1016/j.jnca.2015.11.016.

[19] B. Selvakumar and K. Muneeswaran, "Firefly algorithm based feature selection for network intrusion detection," Comput. Secur., vol. 81, pp. 148–155, 2019, doi: 10.1016/j.cose.2018.11.005.

[20] S. Mohammadi, H. Mirvaziri, M. Ghazizadeh-Ahsaee, and H. Karimipour, "Cyber intrusion detection by combined feature selection algorithm," J. Inf. Secur. Appl., vol. 44, pp. 80–88, 2019, doi: 10.1016/j.jisa.2018.11.007.

[21] F. E. Laghrissi, S. Douzi, K. Douzi, and B. Hssina, "IDS-attention: an efficient algorithm for intrusion detection systems using attention mechanism," J. Big Data, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00544-5.

[22] M. Khudadad and Z. Huang, "Intrusion Detection with Tree-Based Data Mining Classification Techniques by Using KDD," Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng. LNICST, vol. 227, pp. 294–303, 2018.

[23] S. Oshima, T. Nakashima, and Y. Nishikido, "Extraction of characteristics of anomaly accessed IP packets using chi-square method," Proc. Int. Conf. Complex, Intell. Softw. Intensive Syst. CISIS 2009, 2009.

[24] Y. Q. Minchao Ye, Yongqiu Xu, Huijuan Lu, Ke Yan and College, "Cross-Scene Feature Selection for Hyperspectral Images Based on Cross-Domain Information Gain," IGARSS, 2018.

[25] O. Goldstein, M. Kachuee, K. Karkkainen, and M. Sarrafzadeh, "Target-Focused Feature Selection Using Uncertainty Measurements in Healthcare Data," ACM Trans. Comput. Healthc., vol. 1, no. 3, pp. 1–17, May 2020, doi: 10.1145/3383685.

[26] R. P. Priyadarsini and S. Sivakumari, "Gain Ratio Based Feature Selection Method for Privacy Preservation," ICTACT J. SOFT Comput., vol. 01, pp. 201–205, 2011, doi: 10.21917/ijsc.2011.0031.

[27] Z. Li and G. Yan, "A Spark Platform-based Intrusion Detection System by Combining MSMOTE and Improved Adaboost Algorithms," IEEE, 2018, doi: 10.1109/ICSESS.2018.8663723.

[28] N. Farnaaz and M. A. Jabbar, "Random Forest Modeling for Network Intrusion Detection System," 12th Int. Multi-Conference Information, IMCIP, vol. 89, pp. 213 – 217, 2016, doi: https://doi.org/10.1016/j.procs.2016.06.047.

[29] D. Nikolov, I. Kordev, and S. Stefanova, "Concept for network intrusion detection system based on recurrent neural network classifier," Int. Sci. Conf. Electron., 2018.

[30] C. D. H. P. Fuhui Long, "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, pp. 1226–1238, 2015, doi: 10.1109/TPAMI.2005.159.

[31] R. R. Reddy, Y. Ramadevi, and K. V. N. Sunitha, "Effective discriminant function for intrusion detection using SVM," 2016 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2016, 2016.

[32] "ROC curve," Cyber Data Scientist, 2022. https://cyberdatascientist.com/learning/cyber-data-scientist-workflow/evaluation/ (accessed Oct. 12, 2022).