



**Klasifikasi Intrusi Jaringan Menggunakan Random Forest dengan
Penyeimbangan Data Menggunakan CTGAN, CopulaGAN, dan
cWGAN pada Dataset NSL-KDD**

SKRIPSI

Untuk memenuhi persyaratan
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer

Oleh

Muhammad Rio Anrosa

NIM 2011016210031

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

APRIL 2025

SKRIPSI

Klasifikasi Intrusi Jaringan Menggunakan *Random Forest* dengan Penyeimbangan Data Menggunakan CTGAN, CopulaGAN, dan cWGAN pada Dataset NSL-KDD

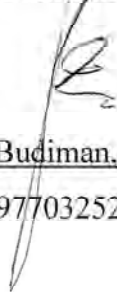
Oleh :

Muhammad Rio Anrosa
NIM. 2011016210031

Telah dipertahankan di depan Dosen Penguji pada tanggal 27 Maret 2025.

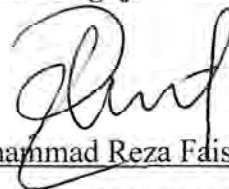
Susunan Dosen Penguji:

Dosen Pembimbing I



Irwan Budiman, S.T., M.Kom.
NIP. 197703252008121001

Dosen Penguji I



Mohammad Reza Faisal, S.T., M.T., Ph.D.
NIP. 197612202008121001

Dosen Pembimbing II



Triando Hamonangan Saragih, S.Kom, M.Kom.
NIP. 199308242019031012


Dosen Penguji II



Radityo Adi Nugroho, S.T., M.Kom.
NIP. 198212042008011006



Banjarbaru, 23 April 2025
Koordinator PS Ilmu Komputer,


Dwi Kartini, S.Kom., M.Kom.
NIP. 198704212012122003

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 25 Maret 2025

Yang Menyatakan,



Muhammad Rio Anrosa

NIM. 2011016210031

ABSTRAK

Klasifikasi Intrusi Jaringan Menggunakan Random Forest dengan Penyeimbangan Data Menggunakan CTGAN, CopulaGAN, dan cWGAN pada Dataset NSL-KDD

(Oleh: Muhammad Rio Anrosa; Pembimbing: Irwan Budiman, S.Kom., M.Kom. dan Triando Hamonangan Saragih. S.Kom., M.Kom.; 2025; 89 halaman)

Intrusi jaringan adalah akses informasi dalam jaringan komputer atau sistem jaringan yang tidak sah dan merugikan. Proses identifikasi intrusi jaringan dapat diterapkan dengan implementasi sistem deteksi intrusi menggunakan metode klasifikasi pembelajaran mesin. Metode klasifikasi yang digunakan pada penelitian ini adalah *random forest classifier*. Dataset yang digunakan pada penelitian ini adalah NSL-KDD dengan 4 kelas data intrusi, DoS, Probe, R2L, dan U2R dengan 1 kelas normal. Dataset ini memiliki ketidakseimbangan distribusi kelas sehingga digunakan arsitektur CTGAN, CopulaGAN, dan cWGAN sebagai penyeimbang data untuk mengatasi masalah tersebut. Data yang diseimbangkan oleh CTGAN, CopulaGAN, dan cWGAN akan diklasifikasikan dan dibandingkan dengan data yang tidak diseimbangkan. Klasifikasi data yang tidak diseimbangkan menghasilkan nilai akurasi sebesar 84,046%, 88,155%, 77,096%, dan 99,325% dan AUC sebesar 84,606%, 72,291%, 50%, dan 50,746%. Klasifikasi data yang diseimbangkan CTGAN menghasilkan nilai akurasi sebesar 87,578%, 93,398%, 84,487%, dan 98,681% serta AUC sebesar 86,613%, 88,232%, 69,719%, dan 78,584%. Selanjutnya Klasifikasi data yang diseimbangkan CopulaGAN menghasilkan nilai akurasi sebesar 88,230%, 91,535%, 80,803%, dan 98,732% serta AUC sebesar 86,7%545, 84,046%, 67,778%, dan 81,575%. Klasifikasi data yang diseimbangkan cWGAN menghasilkan nilai akurasi sebesar 87,129%, 92,392%, 78,398%, dan 98,763% serta AUC sebesar 85,963%, 84,845%, 52,952%, dan 52,686%.

Kata kunci: Intrusi Jaringan, Sistem Deteksi Intrusi, *Random Forest Classifier*, NSL-KDD, Penyeimbang Data, *CTGAN*, *CopulaGAN*, *cWGAN*.

Network Intrusion Classification Using Random Forest with Data Balancing Using CTGAN, CopulaGAN, and cWGAN on NSL-KDD Dataset

(By: Muhammad Rio Anrosa; Supervisor Irwan Budiman, S.Kom., M.Kom. dan Triando Hamonangan Saragih. S.Kom., M.Kom.; 2025; 89 pages)

Network intrusion refers to illegitimate and harmful access to information on a computer network or system. Network intrusion detection can be applied by implementing an intrusion detection system using machine learning classification. The classification method used in this study is the random forest classifier. The dataset used on this study is the NSL-KDD dataset with 4 data intrusion classes, DoS, Probe, R2L, and U2R alongside 1 normal class.. This study uses CTGAN, CopulaGAN, and cWGAN architecture as a data balancer to address the class imbalance issue of NSL-KDD. The imbalanced data and the data balanced by CTGAN, CopulaGAN, dan cWGAN will be classified and compared. The imbalanced data resulted in an accuracy of 84,046%, 88,155%, 77,096%, and 99,325% with AUC scores of 84,606%, 72,291%, 50%, and 50,746. The data balanced using CTGAN resulted in an accuracy 87,578%, 93,398%, 84,487%, and 98,681% with AUC of 86,613%, 88,232%, 69,719%, and 78,584%. Then the data balanced using CopulaGAN gives an accuracy of 88,230%, 91,535%, 80,803%, and 98,732% with AUC scores of 86,7%545, 84,046%, 67,778%, and 81,575%. %. Finally, the data balanced using cWGAN resulted in an accuracy of 87,129%, 92,392%, 78,398%, and 98,763% alongside AUC scores of 85,963%, 84,845%, 52,952%, dan 52,686%.

Keywords: *Network Intrusion, Intrusion Detection System, Random Forest Classifier, NSL-KDD, Data Balancing, CTGAN, CopulaGAN, cWGAN.*

PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh.

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Intrusi Jaringan Menggunakan *Random Forest* dengan Metode *Generative Adversarial Networks* sebagai Penyeimbang Data pada *Dataset* NSL-KDD” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan shalawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga yaumul qiyamah.

Pada lembar ini penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

- (1) Keluarga terutama kepada Orang tua yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini.
- (2) Bapak Irwan Budiman, S.T., M.Kom. selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran penyelesaian skripsi ini.
- (3) Bapak Triando Hamonangan Saragih. S.Kom., M.Kom. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
- (4) Ibu Dwi Kartini, S.Kom., M.Kom. selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
- (5) Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA UNLAM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
- (6) Teman-teman keluarga Ilmu Komputer, yang memberikan bantuan dan dukungan yang senantiasa menjadi alat pacu penulis untuk selalu berusaha dan semangat untuk menyelesaikan skripsi.

(7) Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 25 Maret 2025



Muhammad Rio Anrosa

DAFTAR ISI

BAB I

PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	2
1.3. Tujuan Penelitian.....	3
1.4. Manfaat Penelitian.....	3
TINJAUAN PUSTAKA	4
2.1. Kajian Terdahulu	4
2.2. Landasan Teori	8
2.2.1. Intrusi Jaringan	8
2.2.2. <i>Decision Tree</i>	9
2.2.3. <i>Random Forest</i>	11
2.2.4. <i>Dataset NSL-KDD</i>	12
2.2.5. <i>Generative Adversarial Networks</i>	14
2.2.6. Modifikasi Arsitektur <i>Generative Adversarial Networks</i>	15
2.2.7. <i>Pra-Pemrosesan Data</i>	22
2.2.8. Evaluasi	27
2.3. Batasan Masalah.....	30
METODE PENELITIAN.....	31
3.1. Bahan dan Alat Penelitian	31
3.2. Variabel Penelitian	31
3.3. Prosedur Penelitian	32
3.3.1. Pengumpulan Data.....	33
3.3.2. Pra-pemrosesan Data	33
3.3.3. Penyeimbangan Data	34
3.3.4. Klasifikasi.....	34
3.3.5. Evaluasi	35
HASIL DAN PEMBAHASAN.....	36
4.1. Hasil.....	36
4.1.1. Pengumpulan Data	36
4.1.2. Pra-pemrosesan Data	36
a. Pengolahan <i>Dataset</i>	37
b. Pelabelan Data	39
c. Transformasi Data	41

4.1.3. Penyeimbangan Data	51
a. CTGAN.....	52
b. CopulaGAN	52
c. cWGAN	53
4.1.4. Klasifikasi <i>Random Forest</i>	54
a. Data Tidak Seimbang	54
b. CTGAN.....	59
c. CopulaGAN	63
d. cWGAN	67
4.1.5. Evaluasi	72
4.2. Pembahasan	82
PENUTUP 85	
5.1. Kesimpulan.....	85
5.2. Saran	86
DAFTAR PUSTAKA	87
RIWAYAT PENULIS	109

DAFTAR TABEL

Tabel 1 Keaslian Penelitian.....	6
Tabel 2 Rancangan Penelitian.....	7
Tabel 3 Visualisasi Cara Kerja <i>One-Hot Encoding</i>	23
Tabel 4 <i>Confusion Matrix</i> Kelas Biner	28
Tabel 5 Daftar Fitur NSL-KDD	37
Tabel 6 Potongan Tabel <i>Dataset</i>	39
Tabel 7 Potongan <i>Dataset</i> dengan kolom ‘Intrusion’	41
Tabel 8 Perhitungan Normalisasi fitur ‘Duration’ pada Data Latih.....	42
Tabel 9 Potongan Data Uji setelah Normalisasi	43
Tabel 10 Proses <i>One-Hot Encoding</i> fitur ‘Protocol_type’ pada data latih	43
Tabel 11 Potongan hasil <i>encoding</i> variabel independen.....	44
Tabel 12 Potongan hasil <i>encoding</i> fitur kelas pada data latih.....	44
Tabel 13 Perhitungan Nilai <i>Mean</i> Fitur	45
Tabel 14 Perhitungan Nilai <i>Mean Adjusted Data</i> pada Fitur ‘Duration’	45
Tabel 15 Perhitungan Nilai Kovarian Fitur ‘Duration’ dengan ‘Duration’	46
Tabel 16 Potongan Matriks Kovarian NSL-KDD.....	46
Tabel 17 Nilai <i>Eigen Dataset</i> NSL-KDD	47
Tabel 18 Potongan Vektor <i>Eigen</i> pada <i>Dataset</i> NSL-KDD	47
Tabel 19 Perbandingan Hasil Akurasi Klasifikasi Data PCA.....	49
Tabel 20 Perbandingan Hasil AUC-ROC Klasifikasi Data PCA.....	49
Tabel 21 Potongan <i>Dataframe</i> kelas Normal <i>True</i>	51
Tabel 22 <i>Confusion Matrix</i> Kelas DoS Tanpa Penyeimbangan	55
Tabel 23 Perhitungan Akurasi Kelas DoS Tanpa Penyeimbangan.....	56
Tabel 24 Perhitungan AUC Kelas DoS Tanpa Penyeimbangan	56
Tabel 25 <i>Confusion Matrix</i> Kelas Probe Tanpa Penyeimbangan	56
Tabel 26 Perhitungan Akurasi Kelas Probe Tanpa Penyeimbangan.....	57
Tabel 27 Perhitungan AUC Kelas Probe Tanpa Penyeimbangan.....	57
Tabel 28 <i>Confusion Matrix</i> Kelas R2L Tanpa Penyeimbangan	57
Tabel 29 Perhitungan Akurasi Kelas R2L Tanpa Penyeimbangan.....	58

Tabel 30	Perhitungan AUC Kelas R2L Tanpa Penyeimbangan	58
Tabel 31	<i>Confusion Matrix</i> Kelas U2R Tanpa Penyeimbangan	58
Tabel 32	Perhitungan Akurasi Kelas U2R Tanpa Penyeimbangan	59
Tabel 33	Perhitungan AUC Kelas U2R Tanpa Penyeimbangan.....	59
Tabel 34	<i>Confusion Matrix</i> Kelas DoS yang Diseimbangkan CTGAN	59
Tabel 35	Perhitungan Akurasi Kelas DoS yang Diseimbangkan CTGAN.....	60
Tabel 36	Perhitungan AUC Kelas DoS yang Diseimbangkan CTGAN	60
Tabel 37	<i>Confusion Matrix</i> Kelas Probe yang Diseimbangkan CTGAN	60
Tabel 38	Perhitungan Akurasi Kelas Probe yang Diseimbangkan CTGAN.....	61
Tabel 39	Perhitungan AUC Kelas Probe yang Diseimbangkan CTGAN.....	61
Tabel 40	<i>Confusion Matrix</i> Kelas R2L yang Diseimbangkan CTGAN	61
Tabel 41	Perhitungan Akurasi Kelas R2L yang Diseimbangkan CTGAN.....	62
Tabel 42	Perhitungan AUC Kelas R2L yang Diseimbangkan CTGAN	62
Tabel 43	<i>Confusion Matrix</i> Kelas U2R yang Diseimbangkan CTGAN	62
Tabel 44	Perhitungan Akurasi Kelas U2R yang Diseimbangkan CTGAN	63
Tabel 45	Perhitungan AUC Kelas U2R yang Diseimbangkan CTGAN.....	63
Tabel 46	<i>Confusion Matrix</i> Kelas DoS yang Diseimbangkan CopulaGAN	64
Tabel 47	Perhitungan Akurasi Kelas DoS yang Diseimbangkan CopulaGAN	64
Tabel 48	Perhitungan AUC Kelas DoS yang Diseimbangkan CopulaGAN.....	64
Tabel 49	<i>Confusion Matrix</i> Kelas Probe yang Diseimbangkan CopulaGAN.....	65
Tabel 50	Perhitungan Akurasi Kelas Probe yang Diseimbangkan CopulaGAN..	65
Tabel 51	Perhitungan AUC Kelas Probe yang Diseimbangkan CopulaGAN	65
Tabel 52	<i>Confusion Matrix</i> Kelas R2L yang Diseimbangkan CopulaGAN	66
Tabel 53	Perhitungan Akurasi Kelas R2L yang Diseimbangkan CopulaGAN	66
Tabel 54	Perhitungan AUC Kelas R2L yang Diseimbangkan CopulaGAN.....	66
Tabel 55	<i>Confusion Matrix</i> Kelas U2R yang Diseimbangkan CopulaGAN.....	67
Tabel 56	Perhitungan Akurasi Kelas U2R yang Diseimbangkan CopulaGAN....	67
Tabel 57	Perhitungan AUC Kelas U2R yang Diseimbangkan CopulaGAN.....	67
Tabel 58	<i>Confusion Matrix</i> Kelas DoS yang Diseimbangkan cWGAN	68
Tabel 59	Perhitungan Akurasi Kelas DoS yang Diseimbangkan cWGAN	68
Tabel 60	Perhitungan AUC Kelas DoS yang Diseimbangkan cWGAN.....	68

Tabel 61 <i>Confusion Matrix</i> Kelas Probe yang Diseimbangkan cWGAN.....	69
Tabel 62 Perhitungan Akurasi Kelas Probe yang Diseimbangkan cWGAN	69
Tabel 63 Perhitungan AUC Kelas Probe yang Diseimbangkan cWGAN	69
Tabel 64 <i>Confusion Matrix</i> Kelas R2L yang Diseimbangkan cWGAN	70
Tabel 65 Perhitungan Akurasi Kelas R2L yang Diseimbangkan cWGAN	70
Tabel 66 Perhitungan AUC Kelas R2L yang Diseimbangkan cWGAN.....	70
Tabel 67 <i>Confusion Matrix</i> Kelas U2R yang Diseimbangkan cWGAN.....	71
Tabel 68 Perhitungan Akurasi Kelas U2R yang Diseimbangkan cWGAN.....	71
Tabel 69 Perhitungan AUC Kelas U2R yang Diseimbangkan cWGAN	71
Tabel 70 Hasil Evaluasi Klasifikasi Tanpa <i>Oversampling</i>	72
Tabel 71 Hasil Evaluasi Kelas Biner yang Diseimbangkan CTGAN.....	73
Tabel 72 Hasil Evaluasi Kelas Biner yang Diseimbangkan CopulaGAN	73
Tabel 73 Hasil Evaluasi Kelas Biner yang Diseimbangkan cWGAN	74
Tabel 74 Perbandingan Hasil Evaluasi Kelas DoS	75
Tabel 75 Perbandingan Hasil Evaluasi Kelas Probe	76
Tabel 76 Perbandingan Hasil Evaluasi Kelas R2L	77
Tabel 77 Perbandingan Hasil Evaluasi Kelas U2R.....	79

DAFTAR GAMBAR

Gambar 1 Struktur Umum Decision Tree	9
Gambar 2 Ilustrasi Model <i>Random Forest</i>	12
Gambar 3 Framework G-IDS yang diajukan	15
Gambar 4 Model CTGAN.....	16
Gambar 5 Struktur Generator.....	20
Gambar 6 Struktur Diskriminator	21
Gambar 7 Kurva AUC-ROC.....	29
Gambar 8 Alur Pelaksanaan Penelitian.....	32
Gambar 9 Isi Folder <i>Dataset</i> NSL-KDD	36
Gambar 10 Daftar Serangan untuk Pelabelan Data Intrusi	40
Gambar 11 Modul cWGAN.....	42
Gambar 12 <i>Scree Plot</i> Varians Kumulatif pada 95% Varians	48
Gambar 13 Distribusi Kelas Data Latih dan Data Uji.....	50
Gambar 14 Distribusi Kelas NSL-KDD setelah Data Latih Diseimbangkan	51
Gambar 15 Perbandingan Akurasi pada Kelas DoS	75
Gambar 16 Perbandingan AUC-ROC pada Kelas DoS	76
Gambar 17 Perbandingan Akurasi pada Kelas Probe	77
Gambar 18 Perbandingan AUC-ROC pada Kelas Probe.....	77
Gambar 19 Perbandingan Akurasi pada Kelas R2L	78
Gambar 20 Perbandingan AUC-ROC pada Kelas R2L	78
Gambar 21 Perbandingan Akurasi pada Kelas U2R.....	79
Gambar 22 Perbandingan AUC-ROC pada Kelas U2R.....	80

DAFTAR LAMPIRAN

- Lampiran 1. *Source Code* Label Data
- Lampiran 2. *Source Code* Pelabelan Data
- Lampiran 3. *Source Code* Fungsi Transformasi Data
- Lampiran 4. *Source Code* Transformasi Data
- Lampiran 5. *Source Code* Fungsi Pemetaan Distribusi Kelas
- Lampiran 6. *Source Code Library* Pemodelan
- Lampiran 7. *Source Code* Pemodelan *Random Forest*
- Lampiran 8. *Source Code* Contoh *Testing Dataset*
- Lampiran 9. *Source Code* Pengolahan Data untuk Penyeimbangan
- Lampiran 10. *Source Code* Contoh Penyeimbangan Data Kelas menggunakan CTGAN
- Lampiran 11. *Source Code* Contoh Penyeimbangan Data Kelas menggunakan CopulaGAN
- Lampiran 12. *Source Code* Contoh Penyeimbangan Data Kelas menggunakan cWGAN
- Lampiran 13. *Source Code* Contoh Pembuatan *Dataframe* yang Telah Diseimbangkan