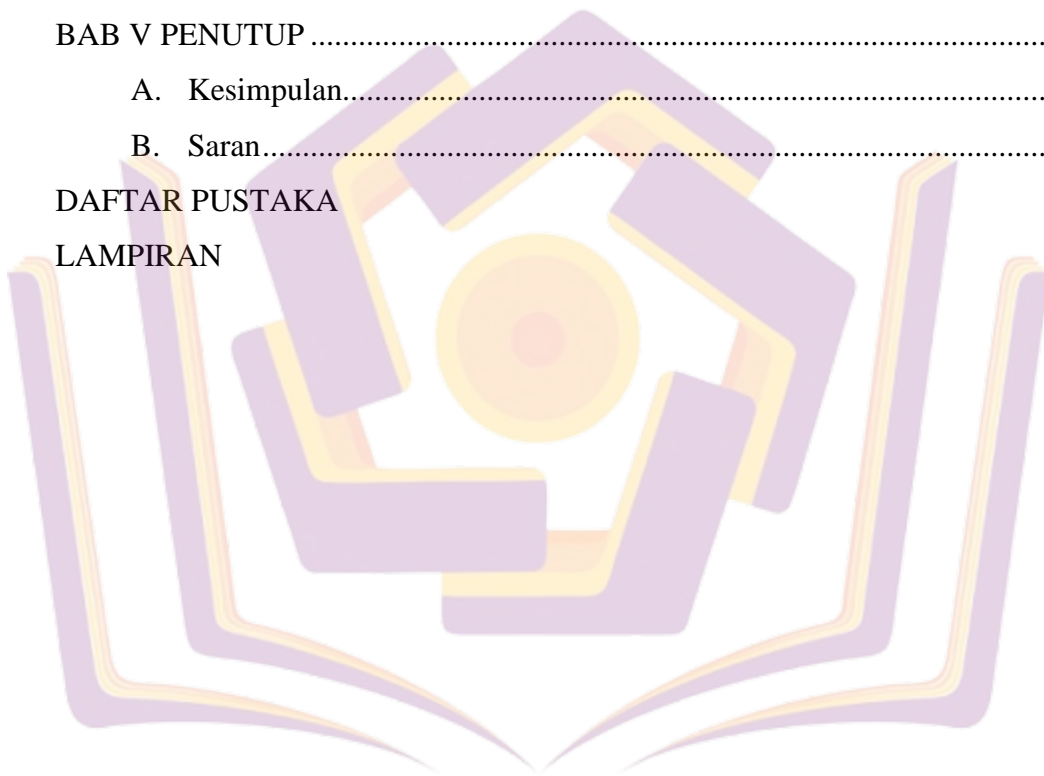


DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN.....	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO.....	viii
KATA PENGANTAR	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR ISTILAH.....	xiv
DAFTAR LAMPIRAN	xx
INTISARI.....	xxi
<i>ABSTRACT</i>	xxii
BAB I PENDAHULUAN.....	1
A. Latar Belakang Masalah.....	1
B. Rumusan Masalah	5
C. Batasan Masalah.....	6
D. Tujuan Penelitian.....	6
E. Manfaat Penelitian.....	7
BAB II TINJAUAN PUSTAKA	8
A. Landasan Teori.....	8
B. Penelitian Sebelumnya	19
BAB III METODE PENELITIAN	24
A. Tempat dan Waktu Penelitian	24
B. Metode Pengumpulan Data	24
C. Alat dan Bahan Penelitian	26
D. Konsep Penelitian.....	27

BAB IV HASIL DAN PEMBAHASAN.....	36
A. Pengumpulan Data	36
B. Analisis Data	37
C. <i>Data Preprocessing</i>	38
D. <i>Feature Selection</i> dengan <i>Random Forest</i>	48
E. <i>Training</i> dan <i>Testing Data</i>	50
F. Pembuatan Model <i>Hybrid CNN</i> dan <i>Random Forest</i>	51
G. Evaluasi Model.....	57
BAB V PENUTUP	67
A. Kesimpulan.....	67
B. Saran.....	68
DAFTAR PUSTAKA	
LAMPIRAN	



DAFTAR TABEL

Tabel 2.1 Tipe Integrasi Model Hybrid.....	9
Tabel 2.2 Kategori Metode Feature selection	11
Tabel 2.3 Jenis Serangan pada dataset UNSW-NB15	12
Tabel 2.4 Confusion matrix.....	17
Tabel 2.5 Penelitian Sebelumnya.....	22
Tabel 3.1 Fitur dataset UNSW-NB15	25
Tabel 3.2 Confusion matrix.....	34
Tabel 4.1 Nilai sebelum dan sesudah pengisian modus.....	40
Tabel 4.2 Nilai sebelum dan sesudah pengisian median.....	41
Tabel 4.3 Hasil sebelum dan sesudah encoding.....	46
Tabel 4.4 Standarisasi fitur numerik	47
Tabel 4.5 Hasil 24 fitur terpilih.....	49
Tabel 4.6 Confusion matrix model hybrid	58
Tabel 4.7 Confusion matrix model random forest	60
Tabel 4.8 Confusion matrix model CNN	62
Tabel 4.9 Perbandingan semua model	65

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Model Random Forest (Khairunnisa, dkk., 2024).....	10
Gambar 2.2 Arsitektur CNN (Ola, dkk., 2025).....	11
Gambar 3.1 Tahapan Penelitian	28
Gambar 4.1 Import Dataset.....	36
Gambar 4.2 Lima baris data teratas	37
Gambar 4.3 Missing values sebelum pemrosesan	39
Gambar 4.4 Missing values sesudah pemrosesan	42
Gambar 4.5 Distribusi kategori serangan.....	44
Gambar 4.6 Fitur numerik dan kategorikal.....	45
Gambar 4.7 Encoding kolom kategorikal	45
Gambar 4.8 Training dan testing data.....	50
Gambar 4.9 Bentuk data CNN	52
Gambar 4.10 Arsitektur model CNN	53
Gambar 4.11 Epoch 1 sampai 3 model CNN.....	53
Gambar 4.12 Epoch 4 sampai 50 model CNN.....	55
Gambar 4.13 Bentuk ekstraksi fitur CNN.....	56
Gambar 4.14 Bentuk penggabungan fitur	56
Gambar 4.15 Training random forest pada fitur hybrid.....	57
Gambar 4.16 Pohon Keputusan Random Forest	64

DAFTAR ISTILAH

1. Accuracy (Akurasi): Metrik evaluasi yang mengukur proporsi prediksi benar (baik positif maupun negatif) terhadap keseluruhan data yang diuji.
2. Anomaly-based Detection: Metode deteksi intrusi yang mengidentifikasi ancaman berdasarkan penyimpangan dari pola perilaku normal sistem.
3. Artificial Neural Network: Sistem komputasi yang terinspirasi dari cara kerja otak manusia, terdiri dari neuron-neuron buatan yang saling terhubung untuk memproses informasi.
4. Autoencoder: Arsitektur deep learning yang digunakan untuk pembelajaran representasi data secara tidak terawasi dengan mengompresi dan merekonstruksi input.
5. Backdoor: Pintu masuk tersembunyi yang ditanamkan penyerang pada sistem untuk memperoleh akses tidak sah tanpa melalui mekanisme otentikasi yang sah.
6. Binary Classification: Tugas klasifikasi yang membedakan data ke dalam dua kategori, dalam penelitian ini membedakan traffic normal dan serangan.
7. Bootstrapping: Teknik sampling with replacement yang digunakan dalam Random Forest untuk membuat subset data pelatihan yang berbeda untuk setiap pohon keputusan.
8. Cascaded Hybrid: Tipe model hybrid yang menyusun model secara bertingkat dengan output model pertama menjadi input model kedua.
9. CIA Triad: Prinsip dasar keamanan informasi yang terdiri dari Confidentiality (kerahasiaan), Integrity (integritas), dan Availability (ketersediaan).
10. Confusion Matrix: Tabel yang merangkum performa model klasifikasi dengan menampilkan jumlah prediksi benar dan salah untuk setiap kelas.
11. Convolutional Layer: Lapisan dalam CNN yang menerapkan operasi konvolusi menggunakan filter untuk mengekstraksi fitur dari data input.
12. Convolutional Neural Network (CNN): Arsitektur deep learning yang dirancang untuk memproses data dengan struktur grid seperti gambar, menggunakan lapisan konvolusi untuk ekstraksi fitur otomatis.

13. Cumulative Importance: Akumulasi nilai feature importance secara berurutan untuk menentukan threshold pemilihan fitur.
14. Data Cleaning: Proses pembersihan data yang mencakup penanganan missing values, duplikasi, dan inkonsistensi untuk meningkatkan kualitas data.
15. Data Preprocessing: Tahapan persiapan data mentah menjadi format yang sesuai untuk analisis dan pemodelan machine learning.
16. Decision Level Hybrid: Tipe model hybrid yang menggabungkan prediksi dari beberapa model menggunakan voting, rata-rata, atau meta-classifier.
17. Decision Tree: Algoritma supervised learning yang membuat keputusan dengan memecah data menjadi subset berdasarkan nilai fitur tertentu dalam struktur pohon.
18. Deep Learning: Cabang machine learning yang menggunakan artificial neural networks dengan banyak lapisan untuk mempelajari representasi data yang kompleks dan hierarkis.
19. Denial of Service (DoS): Serangan yang membanjiri sistem target dengan lalu lintas berlebihan sehingga layanan menjadi tidak dapat diakses oleh pengguna sah.
20. Distributed Denial of Service (DDoS): Serangan DoS yang dilakukan secara terdistribusi dari banyak sumber secara bersamaan untuk melumpuhkan target.
21. Embedding Methods: Kategori feature selection yang melakukan seleksi fitur secara otomatis selama proses pelatihan model.
22. Ensemble Learning: Teknik machine learning yang menggabungkan prediksi dari beberapa model untuk meningkatkan akurasi dan stabilitas prediksi.
23. Epoch: Satu iterasi lengkap dari proses pelatihan model neural network di mana seluruh dataset training dilewatkan melalui jaringan.
24. Exploits: Teknik serangan yang memanfaatkan kelemahan atau celah keamanan pada perangkat lunak atau sistem operasi untuk mendapatkan akses ilegal.
25. Exploratory Data Analysis (EDA): Proses analisis data untuk memahami karakteristik, pola, distribusi, dan struktur data melalui visualisasi dan statistik deskriptif.

26. F1-Score: Metrik evaluasi yang merupakan harmonic mean dari presisi dan recall, memberikan keseimbangan antara kedua metrik tersebut.
27. False Negative (FN): Kondisi di mana model salah memprediksi kelas negatif padahal sebenarnya positif (serangan diprediksi sebagai normal).
28. False Positive (FP): Kondisi di mana model salah memprediksi kelas positif padahal sebenarnya negatif (normal diprediksi sebagai serangan).
29. Feature Extractor: Komponen model yang bertugas mengekstraksi representasi fitur dari data input, dalam penelitian ini menggunakan CNN.
30. Feature Fusion: Proses penggabungan fitur dari berbagai sumber atau representasi untuk meningkatkan kemampuan prediktif model.
31. Feature Importance: Nilai yang mengukur kontribusi atau kepentingan setiap fitur terhadap performa model dalam melakukan prediksi.
32. Feature Level Hybrid: Tipe model hybrid yang menggabungkan fitur dari beberapa metode sebelum dimasukkan ke model klasifikasi utama.
33. Feature Selection: Proses memilih subset fitur yang paling relevan dari dataset untuk mengurangi dimensi dan meningkatkan performa model.
34. Filter Methods: Kategori feature selection yang menggunakan metrik statistik untuk mengevaluasi fitur secara independen dari model.
35. Fully Connected Layer: Lapisan dalam neural network di mana setiap neuron terhubung dengan semua neuron pada lapisan sebelumnya.
36. Fuzzers: Teknik pengujian keamanan yang mengirimkan input acak atau tidak valid ke sistem untuk menemukan kelemahan atau bug.
37. Gradient Descent: Algoritma optimasi yang digunakan untuk meminimalkan fungsi loss dengan melakukan iterasi pembaruan parameter model.
38. Hierarchical Feature Learning: Pembelajaran fitur secara berlapis di mana setiap lapisan mengekstrak representasi yang semakin abstrak.
39. Hyperparameter Tuning: Proses optimasi parameter model yang tidak dipelajari dari data, seperti learning rate, jumlah layer, atau jumlah estimator.
40. Imbalanced Dataset: Kondisi dataset di mana jumlah sampel antar kelas tidak seimbang, seperti traffic normal yang mendominasi dibanding serangan.

41. Intrusion Detection System (IDS): Sistem keamanan yang dirancang untuk memantau aktivitas jaringan dan mengidentifikasi potensi ancaman atau serangan.
42. Label Encoding: Teknik preprocessing yang mengubah data kategorikal menjadi representasi numerik dengan memetakan setiap kategori ke angka integer.
43. Local Minima: Titik pada fungsi loss di mana nilai loss lebih rendah dari titik-titik sekitarnya namun bukan nilai minimum global.
44. Loss Function: Fungsi yang mengukur perbedaan antara prediksi model dengan nilai aktual, digunakan untuk mengoptimasi parameter model.
45. Machine Learning: Cabang kecerdasan buatan yang memungkinkan sistem belajar dari data dan membuat prediksi tanpa diprogram secara eksplisit.
46. Max Pooling: Operasi dalam CNN yang mengurangi dimensi spatial dengan mengambil nilai maksimum dari region tertentu pada feature map.
47. Median: Nilai tengah dalam dataset terurut, digunakan untuk mengisi missing values pada data numerik karena lebih robust terhadap outlier.
48. Missing Values: Nilai yang hilang atau tidak tersedia pada dataset yang perlu ditangani sebelum melakukan analisis atau pemodelan.
49. Model Hybrid: Pendekatan yang menggabungkan dua atau lebih teknik, arsitektur, atau algoritma untuk memanfaatkan kekuatan masing-masing komponen.
50. Modus: Nilai yang paling sering muncul dalam dataset, digunakan untuk mengisi missing values pada data kategorikal.
51. Normalization: Proses penskalaan nilai fitur ke dalam rentang tertentu, biasanya $[0,1]$, untuk memastikan semua fitur memiliki skala yang setara.
52. Overfitting: Kondisi di mana model terlalu menyesuaikan dengan data training sehingga performa pada data testing menurun.
53. Pooling Layer: Lapisan dalam CNN yang mengurangi dimensi spatial dari feature maps sambil mempertahankan informasi penting.
54. Port Scanning: Aktivitas pengintaian untuk memetakan port yang terbuka pada sistem target dan mengidentifikasi layanan yang berjalan.

55. Precision (Presisi): Metrik evaluasi yang mengukur proporsi prediksi positif yang benar terhadap seluruh prediksi positif.
56. Random Forest: Algoritma ensemble learning yang membangun banyak pohon keputusan dan menggabungkan hasilnya untuk meningkatkan akurasi dan mengurangi overfitting.
57. Recall: Metrik evaluasi yang mengukur proporsi data positif aktual yang berhasil diidentifikasi dengan benar oleh model.
58. Reconnaissance: Aktivitas pengintaian awal yang dilakukan penyerang untuk mengumpulkan informasi tentang target sebelum melakukan serangan.
59. Recurrent Neural Network (RNN): Arsitektur neural network yang dirancang untuk memproses data sekuensial dengan mempertahankan informasi dari input sebelumnya.
60. Representation Level Hybrid: Tipe model hybrid di mana model pertama menghasilkan representasi fitur yang digunakan sebagai input untuk model kedua.
61. Reshaping: Proses mengubah dimensi array atau tensor tanpa mengubah datanya untuk menyesuaikan dengan format input yang dibutuhkan model.
62. Shellcode: Kode berbahaya yang dieksekusi untuk mengambil alih kendali sistem korban, biasanya digunakan dalam serangan exploit.
63. Signature-based Detection: Metode deteksi intrusi yang mencocokkan pola serangan dengan database tanda tangan serangan yang sudah dikenal.
64. Standardization: Proses penskalaan fitur sehingga memiliki mean 0 dan standar deviasi 1 menggunakan z-score normalization.
65. Stratified Split: Metode pembagian data yang mempertahankan proporsi kelas pada data training dan testing agar representatif.
66. Supervised Learning: Paradigma machine learning yang melatih model menggunakan data berlabel untuk mempelajari hubungan antara input dan output.
67. Training Set: Subset data yang digunakan untuk melatih model machine learning agar dapat mempelajari pola dari data.

68. Testing Set: Subset data yang digunakan untuk mengevaluasi performa model pada data yang belum pernah dilihat sebelumnya.
69. True Negative (TN): Kondisi di mana model dengan benar memprediksi kelas negatif (normal diprediksi sebagai normal).
70. True Positive (TP): Kondisi di mana model dengan benar memprediksi kelas positif (serangan diprediksi sebagai serangan).
71. UNSW-NB15: Dataset serangan jaringan yang dikembangkan oleh Australian Centre for Cyber Security, berisi lalu lintas jaringan normal dan sembilan jenis serangan modern.
72. Validation Set: Subset data yang digunakan untuk mengevaluasi performa model selama proses training dan tuning hyperparameter.
73. Worms: Malware yang dapat mereplikasi diri secara otomatis dan menyebar ke seluruh jaringan tanpa interaksi pengguna.
74. Wrapper Methods: Kategori feature selection yang mengevaluasi subset fitur berdasarkan performa model secara langsung.
75. Z-score Normalization: Teknik standarisasi yang mentransformasi data menjadi distribusi dengan mean 0 dan standar deviasi 1.

DAFTAR LAMPIRAN

Lampiran 1. Kartu Bimbingan Dosen Pembimbing 1

Lampiran 2. Kartu Bimbingan Dosen Pembimbing 2

Lampiran 3. Link Dataset UNSW-NB15

Lampiran 4. Kode Program

