REDTECH PUTRA BENUA

Bagaimana AI Mengubah Cara Kita Melawan Ancaman Siber

Konsep, Model dan Implementasi

PAULUS HARSADI
YOVITA KINANTI KUMARAHADI
HENDRO WIJAYANTO
DWI REMAWATI

Bagaimana AI Mengubah Cara Kita Melawan Ancaman Siber

Konsep, Model, dan Implementasi

Penulis :

Paulus Harsadi

Yovita Kinanti Kumarahadi

Hendro Wijayanto

Dwi Remawati

Penerbit
REDTECH PUTRA BENUA

Bagaimana AI Mengubah Cara Kita Melawan Ancaman Siber. Konsep, Model, dan Implementasi

Penulis:
Paulus Harsadi, Yovita Kinanti Kumarahadi, Hendro Wijayanto, Dwi Remawati

Ucapan terima kasih disampaikan kepada Kementerian Pendidikan Tinggi, Sains, dan Teknologi Republik Indonesia atas dukungan finansial melalui Hibah Penelitian Dosen Pemula Tahun Anggaran 2026 dengan judul Intelligent Cybersecurity Mitigation Recommendation System: Pendekatan Adaptif dengan Graph Neural Networks. Dan Nomor Kontrak 285/C3/DT.05.00/PL-BARU/2026 tanggal 30 Januari 2026 dan No. Kontrak Turunan 071/LL6/AL.04.03/PL-BARU/2026 tanggal 2 Februari 2026

e-ISBN:
978-634-05-3679-9

Cetakan Pertama:
Juli 2026

Penerbit REDTECH PUTRA BENUA

Redtech Putra Benua
(Grup CV. Redtech Putra Benua)
Perumahan Graha Jatikuwung Indah Blok D04
Wonosari, RT.002 RW.003 Jatikuwung, Gondangrejo, Karanganyar, 57188

Website: www.redtechidn.org
Instagram: @redtechidn
Whatsapp : +6281613 55055
Homepage : https://redtechidn.org

Daftar Isi

Kata Pengantar

Puji dan syukur penulis panjatkan atas selesainya penulisan buku ini. Kehadiran buku ini berangkat dari sebuah kegelisahan sederhana: di satu sisi, ancaman siber di sekitar kita bergerak semakin cepat, semakin cerdas, dan semakin sulit diprediksi dengan pendekatan konvensional; di sisi lain, literatur berbahasa Indonesia yang membahas irisan antara kecerdasan buatan dan keamanan siber secara utuh mulai dari fondasi matematis hingga implementasi nyata di lapangan masih sangat terbatas. Buku ini disusun untuk menjawab kesenjangan tersebut.

Judul **Bagaimana AI Mengubah Cara Kita Melawan Ancaman Siber** dipilih bukan sekadar untuk menarik perhatian, melainkan karena judul ini secara jujur menggambarkan apa yang ingin disampaikan buku ini bahwa kecerdasan buatan bukan sekadar tren teknologi yang lewat begitu saja, melainkan pergeseran paradigma yang nyata dalam cara kita mendeteksi, memahami, dan merespons ancaman digital. Pergeseran ini tidak terjadi dalam ruang hampa akademik semata, tetapi juga di ruang server, di ruang rapat pengambil kebijakan, dan di layar-layar security operations center yang bekerja tanpa henti menjaga infrastruktur digital kita.

Buku ini disusun dalam empat bagian besar. Bagian I mengajak pembaca memahami mengapa pendekatan keamanan siber konvensional tidak lagi memadai di hadapan ancaman modern. Bagian II membangun fondasi teknis secara bertahap, dari machine learning klasik hingga graph neural network dan large language model. Bagian III menunjukkan bagaimana fondasi tersebut diwujudkan menjadi sistem nyata, termasuk studi kasus implementasi yang menjadi jantung buku ini. Bagian IV menutup dengan refleksi kritis atas tantangan etis, kolaborasi lintas organisasi,

tantangan nyata di industri, serta arah masa depan menuju pertahanan siber yang otonom.

Penulis menyampaikan terima kasih yang mendalam kepada instansi-instansi yang telah memberikan ruang dan dukungan bagi penulis untuk terus berkarya, kepada rekan-rekan sejawat yang telah memberikan masukan berharga selama proses penulisan, serta kepada keluarga yang senantiasa memberikan dukungan dan pengertian di setiap tahap penyelesaian buku ini. Penulis juga berterima kasih kepada komunitas riset keamanan siber dan kecerdasan buatan, baik nasional maupun internasional, yang karya-karyanya menjadi rujukan penting dalam penyusunan buku ini.

Akhir kata, penulis berharap buku ini dapat memberikan kontribusi yang bermakna bagi penguatan ketahanan digital masyarakat dan bangsa Indonesia, serta menjadi bekal bagi generasi berikutnya dalam menghadapi ancaman siber yang terus berevolusi. Selamat membaca.

Surakarta, Juli 2026
Penulis

BAB 1

Lanskap Ancaman Siber di Era Transformasi Digital

1.1 Transformasi Digital dan Perluasan Permukaan Serangan

Transformasi digital telah mengubah cara individu, organisasi, dan negara beroperasi. Adopsi cloud computing, Internet of Things (IoT), sistem kerja hybrid, layanan finansial digital, serta ketergantungan pada rantai pasok teknologi informasi yang saling terhubung telah memperluas apa yang dalam kajian keamanan siber disebut sebagai attack surface, yaitu keseluruhan titik yang berpotensi dieksploitasi oleh pihak yang tidak berwenang. Setiap perangkat baru yang terhubung, setiap layanan berbasis application programming interface (API), dan setiap proses bisnis yang dipindahkan ke ranah digital menambah kompleksitas yang harus diamankan.

Perluasan permukaan serangan ini tidak diimbangi secara merata dengan penguatan tata kelola keamanan. Banyak organisasi, khususnya usaha kecil dan menengah, mengadopsi teknologi digital lebih cepat daripada kemampuan mereka membangun kapasitas keamanan siber yang memadai. Kesenjangan inilah yang kemudian dieksploitasi oleh pelaku ancaman, baik individu, kelompok kriminal terorganisir, maupun aktor yang disponsori negara.

1.2 Evolusi Ancaman Siber: Dari Pendekatan Konvensional Menuju Serangan Bertenaga Kecerdasan Buatan

Ancaman siber telah melalui beberapa fase evolusi. Pada fase awal, serangan umumnya bersifat oportunistik dan menggunakan malware dengan tanda tangan (signature) yang relatif statis, sehingga cukup efektif dideteksi oleh sistem keamanan berbasis basis data ancaman yang dikenal. Fase berikutnya ditandai dengan munculnya advanced persistent threat (APT), yaitu serangan yang terencana, bertahan lama dalam sistem target, dan menyasar entitas bernilai tinggi seperti infrastruktur kritis atau lembaga pemerintah.

Fase terkini, yang menjadi fokus pembahasan buku ini, ditandai oleh masuknya kecerdasan buatan sebagai alat bantu utama pelaku serangan. Laporan FortiGuard Labs 2026 mencatat bahwa kejahatan siber kini tidak lagi beroperasi sebagai kampanye-kampanye yang terisolasi, melainkan sebagai suatu sistem yang terindustrialisasi dan sangat otomatis, dengan pelaku memanfaatkan "shadow agents" berbasis AI untuk mempersingkat keseluruhan siklus hidup serangan, mulai dari pengintaian hingga eksploitasi (FortiGuard Labs, 2026). Sejalan dengan itu, laporan Rapid7 Labs 2026 menegaskan bahwa jendela waktu prediktif dalam keamanan siber telah menyempit drastis: eksploitasi kerentanan kini terjadi lebih cepat daripada siklus remediasi mampu merespons (Rapid7 Labs, 2026).

Penting untuk dicatat bahwa peran AI dalam lanskap ancaman ini, menurut temuan Rapid7, lebih banyak berfungsi mempercepat dan memperbesar skala taktik serangan yang sudah dikenal, ketimbang menciptakan jenis serangan yang sepenuhnya baru. AI mempercepat proses pengintaian, mempercanggih dan mempercepat penyusunan phishing, serta

memampatkan garis waktu serangan sehingga jarak antara eksposur dan dampak semakin singkat (Rapid7 Labs, 2026).

1.3 Karakteristik Serangan Siber Modern

Beberapa karakteristik menonjol yang membedakan ancaman siber masa kini dari ancaman satu dekade lalu meliputi:

1. **Kecepatan (velocity).** Risiko siber saat ini tidak lagi ditentukan semata oleh kecanggihan teknik serangan, melainkan oleh kecepatan eksekusinya. Kerentanan yang baru diungkap dapat dieksploitasi dalam hitungan jam, bukan lagi hitungan minggu, sehingga jendela waktu bagi tim pertahanan untuk merespons nyaris tidak tersisa (FortiGuard Labs, 2026).
2. **Otomatisasi dan industrialisasi.** Ekosistem kejahatan siber kini menyerupai rantai pasok industri, lengkap dengan pembagian peran, model layanan seperti ransomware-as-a-service, serta pemanfaatan agen otomatis untuk menjalankan tahapan serangan secara paralel (FortiGuard Labs, 2026).
3. **Adaptivitas dan personalisasi.** Model AI generatif memungkinkan pelaku menyusun pesan phishing, dokumen palsu, maupun suara dan video deepfake yang sangat meyakinkan serta disesuaikan dengan profil korban, sehingga mampu melewati mekanisme deteksi konvensional maupun kewaspadaan manusia (Vectra AI, 2026; World Economic Forum, 2026).
4. **Pergeseran motif ke arah penipuan finansial.** Untuk pertama kalinya, penipuan bertenaga siber (cyber-enabled fraud) menggeser ransomware sebagai kekhawatiran utama jajaran eksekutif secara global, seiring meningkatnya insiden pencurian identitas, business

email compromise, dan penipuan berbasis deepfake (World Economic Forum & Accenture, 2026).

5. **Fragmentasi pertahanan sebagai celah utama.** Penggunaan berbagai perangkat dan platform keamanan yang tidak terintegrasi menciptakan celah deteksi yang memperlambat respons insiden, sehingga mendorong pergeseran paradigma menuju pendekatan zero trust dan verifikasi identitas berkelanjutan (Constellation Research, 2025).

1.4 Dampak terhadap Kerahasiaan, Integritas, dan Ketersediaan Sistem Digital

Secara konseptual, dampak ancaman siber dapat dipetakan pada tiga pilar keamanan informasi yang dikenal sebagai triad CIA: **Confidentiality** (kerahasiaan), **Integrity** (integritas), dan **Availability** (ketersediaan).

- Kerahasiaan terancam ketika data sensitif baik data pribadi, data finansial, maupun kekayaan intelektual diakses atau dibocorkan oleh pihak yang tidak berwenang, termasuk melalui kebocoran yang terjadi akibat penyalahgunaan perangkat AI generatif di lingkungan kerja.
- Integritas terancam ketika data atau sistem dimanipulasi tanpa izin, misalnya melalui malware yang mengubah catatan transaksi, atau melalui konten deepfake yang memalsukan identitas dan komunikasi resmi untuk mengelabui proses verifikasi.
- Ketersediaan terancam ketika serangan seperti ransomware atau distributed denial-of-service (DDoS) melumpuhkan akses terhadap sistem dan layanan penting, sehingga mengganggu kelangsungan operasional organisasi maupun layanan publik.

Ketiga pilar tersebut saling berkaitan. Sebagai contoh, sebuah insiden ransomware modern umumnya tidak hanya mengenkripsi data

(mengancam ketersediaan), tetapi juga mengeksfiltrasi data terlebih dahulu sebagai leverage tambahan (mengancam kerahasiaan), sekaligus berpotensi memanipulasi log sistem untuk menghambat investigasi forensik (mengancam integritas). Karakteristik serangan multidimensi inilah yang menuntut pendekatan pertahanan yang juga bersifat holistik dan adaptif, sebuah kebutuhan yang menjadi dasar argumentasi bagi pendekatan berbasis kecerdasan buatan yang dibahas pada bab-bab selanjutnya.

1.5 Sorotan Data Global Terkini (2025–2026)

Berbagai lembaga riset keamanan siber global secara konsisten melaporkan tren peningkatan volume dan kompleksitas serangan pada periode 2025–2026. Tabel 1.1 merangkum sejumlah temuan kunci yang relevan dengan pembahasan buku ini.

Tabel 1.1 Indikator Lanskap Ancaman Siber Global 2025–2026

Indikator	Nilai / Temuan	Sumber
Kenaikan korban ransomware secara global (YoY)	389%	FortiGuard Labs, 2026 (FortiGuard Labs, 2026)
Kenaikan eksploitasi kerentanan CVSS 7–10 yang baru diungkap (YoY)	105%	Rapid7 Labs, 2026 (Rapid7 Labs, 2026)
Insiden yang didorong oleh akun valid tanpa MFA kuat	43,9% dari investigasi insiden	Rapid7 Labs, 2026 (Rapid7 Labs, 2026)
Keterlibatan ransomware dalam investigasi MDR	42%	Rapid7 Labs, 2026 (Rapid7 Labs, 2026)

Organisasi yang terdampak langsung oleh penipuan bertenaga siber pada 2025	73%	WEF Global Cybersecurity Outlook 2026 (World Economic Forum & Accenture, 2026)
Pemimpin global yang menilai AI sebagai pendorong utama perubahan keamanan siber	94%	WEF Global Cybersecurity Outlook 2026 (World Economic Forum & Accenture, 2026)
Peningkatan rasio klik pada phishing hasil AI dibanding buatan manusia	±4 kali lipat	Vectra AI, 2026 (Vectra AI, 2026)

1.6 Lanskap Ancaman Siber di Indonesia

Sebagai salah satu negara dengan populasi pengguna internet terbesar di dunia dan laju digitalisasi yang tinggi, Indonesia turut menghadapi tekanan ancaman siber yang meningkat tajam. Badan Siber dan Sandi Negara (BSSN) secara rutin merilis data pemantauan anomali trafik nasional yang menggambarkan skala persoalan ini secara empiris. Tabel 1.2 merangkum sejumlah temuan utama.

Tabel 1.2 Indikator Lanskap Ancaman Siber Indonesia 2025–2026

Indikator	Nilai / Temuan	Sumber
Total anomali/serangan siber sepanjang 2025	5,5 miliar (naik 714% dari rata-rata 2020–2024)	BSSN, dikutip Kompas 2026 (BSSN, 2026)

Total anomali serangan, 1 Jan–15 Apr 2026	1,52 miliar	BSSN, dikutip Kompas 2026 (BSSN, 2026)
Total anomali serangan, Januari–Juli 2025	3,64 miliar	BSSN, dikutip Tempo 2025 (BSSN, 2025a)
Proporsi anomali berbasis malware (H1 2025)	83,68%	BSSN, dikutip Tempo 2025 (BSSN, 2025a)
Jenis malware terbanyak terdeteksi 2025	Mirai Botnet, Remcos RAT, Generic Trojan	BSSN, dikutip Bisnis.com 2025 (BSSN, 2025b)
Kerugian ekonomi akibat kejahatan siber 2024	±Rp18 triliun (naik 30% dari tahun sebelumnya)	BSSN, dikutip Verihubs 2026 (BSSN & Kemkomdigi, 2026)
Kenaikan volume konten deepfake di Indonesia (5 tahun terakhir)	550%	Kemkomdigi, dikutip Verihubs 2026 (BSSN & Kemkomdigi, 2026)

Data tersebut menunjukkan bahwa ancaman siber di Indonesia bukan lagi sekadar potensi risiko, melainkan realitas operasional yang dihadapi setiap hari oleh penyelenggara sistem elektronik, baik di sektor pemerintahan, keuangan, maupun usaha kecil dan menengah. Dominasi malware yang adaptif, ditambah dengan meningkatnya pemanfaatan AI oleh pelaku kejahatan digital, mendorong pihak berwenang menegaskan bahwa pendekatan keamanan konvensional yang mengandalkan basis data ancaman yang telah dikenal tidak lagi memadai (BSSN, 2026). Kondisi ini turut mendorong percepatan pembahasan Rancangan Undang-Undang Keamanan dan Ketahanan Siber (RUU KKS) sebagai landasan hukum yang lebih kuat bagi tata kelola keamanan siber nasional (BSSN, 2025a).

Catatan: Fenomena Ancaman Siber Terbaru (2025–2026)

- *Serangan bertenaga agentic AI ("shadow agents") mulai digunakan untuk memampatkan siklus hidup serangan, dari pengintaian hingga eksploitasi, tanpa banyak campur tangan manusia (FortiGuard Labs, 2026).*
- *Ransomware-as-a-Service (RaaS) semakin terindustrialisasi; jumlah korban ransomware global tercatat melonjak 389% secara tahunan pada laporan 2026 (FortiGuard Labs, 2026).*
- *Jendela waktu antara pengungkapan kerentanan dan eksploitasi aktif menyusut drastis dari hitungan minggu menjadi hitungan jam, sehingga siklus tambal (patch cycle) konvensional tertinggal (FortiGuard Labs, 2026; Rapid7 Labs, 2026).*
- *Penipuan bertenaga siber (cyber-enabled fraud), termasuk deepfake audio/video, untuk pertama kalinya melampaui ransomware sebagai kekhawatiran utama CEO secara global menjelang 2026 (World Economic Forum & Accenture, 2026).*
- *Satu insiden penipuan deepfake video call tercatat merugikan sebuah perusahaan rekayasa global hingga USD 25,6 juta dalam satu kejadian (Vectra AI, 2026).*
- *Email phishing hasil rekayasa AI generatif tercatat memiliki tingkat klik hingga empat kali lebih tinggi dibandingkan phishing buatan manusia (Vectra AI, 2026).*
- *Di Indonesia, BSSN mencatat lonjakan serangan siber tujuh kali lipat pada 2025 dibandingkan rata-rata lima tahun sebelumnya, dengan tren peningkatan yang berlanjut pada awal 2026 dan didominasi oleh malware yang semakin adaptif (BSSN, 2026; BSSN, 2025a).*
- *Volume konten deepfake di Indonesia meningkat 550% dalam lima tahun terakhir, dengan pemanfaatan yang mulai mengarah pada*

pemalsuan identitas saat proses verifikasi digital di sektor perbankan dan fintech (BSSN & Kemkomdigi, 2026).

1.7 Implikasi bagi Pendekatan Keamanan Berbasis Kecerdasan Buatan

Pergeseran karakteristik ancaman siber dari serangan yang lambat dan berpola tetap menuju serangan yang cepat, otomatis, dan adaptif membawa implikasi mendasar bagi strategi pertahanan. Pendekatan keamanan berbasis aturan statis (rule-based) dan basis data tanda tangan ancaman yang telah dikenal (signature-based) semakin kesulitan mengimbangi laju perubahan taktik pelaku serangan. Sebagaimana ditegaskan oleh berbagai pemangku kepentingan keamanan siber, ancaman yang didukung AI menuntut pendekatan pertahanan yang juga memanfaatkan teknologi serupa: sistem yang mampu belajar dari pola data, beradaptasi terhadap ancaman baru, dan memberikan respons yang kontekstual (BSSN, 2026).

Kebutuhan inilah yang melatarbelakangi pembahasan kecerdasan buatan sebagai pendekatan pertahanan siber pada bab-bab selanjutnya mulai dari fondasi teknis machine learning dan deep learning, pendekatan berbasis graf seperti Graph Neural Networks untuk memodelkan hubungan antarentitas dalam jaringan, hingga studi kasus implementasi sistem rekomendasi mitigasi yang adaptif dan kontekstual.

1.8 Rangkuman

Transformasi digital memperluas permukaan serangan (attack surface) organisasi secara signifikan, tanpa selalu diimbangi penguatan tata kelola keamanan yang setara. Ancaman siber telah berevolusi dari pola serangan statis menuju ekosistem kejahatan yang terindustrialisasi, otomatis, dan dipercepat oleh kecerdasan buatan. Karakteristik serangan modern ditandai oleh kecepatan eksploitasi yang tinggi, otomatisasi berskala industri, personalisasi berbasis AI generatif, serta pergeseran motif ke arah penipuan finansial. Dampak ancaman siber dapat dipetakan pada tiga pilar keamanan informasi, kerahasiaan, integritas, dan ketersediaan, yang saling berkaitan dalam insiden serangan modern. Data global maupun nasional (Indonesia) pada periode 2025–2026 secara konsisten menunjukkan tren peningkatan volume dan kompleksitas serangan, memperkuat urgensi pendekatan pertahanan berbasis kecerdasan buatan.

Daftar Pustaka

FortiGuard Labs. (2026). 2026 Global Threat Landscape Report. Fortinet, Inc.

Rapid7 Labs. (2026). 2026 Global Threat Landscape Report. Rapid7, Inc.

World Economic Forum & Accenture. (2026). Global Cybersecurity Outlook 2026. World Economic Forum.

Vectra AI. (2026). AI Scams in 2026: How They Work and How to Detect Them.

Badan Siber dan Sandi Negara (BSSN). (2026). Data lonjakan serangan siber Indonesia 2025–2026, dikutip dalam Kompas.com, 23 April 2026.

Badan Siber dan Sandi Negara (BSSN). (2025a). Data anomali serangan siber Januari–Juli 2025, dikutip dalam Tempo.co, 8 Agustus 2025.

Badan Siber dan Sandi Negara (BSSN). (2025b). Data anomali trafik hingga September 2025, dikutip dalam Bisnis.com, 1 Desember 2025.

Badan Siber dan Sandi Negara (BSSN) & Kementerian Komunikasi dan Digital (Kemkomdigi). (2026). Statistik cybercrime dan deepfake Indonesia, dikutip dalam Verihubs, 22 Maret 2026.

World Economic Forum. (2026). The Trends Reshaping Cybersecurity. Dalam Global Cybersecurity Outlook 2026.

Constellation Research. (2025). Analisis fragmentasi sistem keamanan dan pergeseran menuju zero trust, dikutip dalam Beritajejakfakta.id, 14 April 2026.

BAB 2

Mengapa Kecerdasan Buatan? Keterbatasan Pendekatan Keamanan Siber Konvensional

2.1 Pendekatan Keamanan Siber Konvensional: Tinjauan Ringkas

Selama beberapa dekade, pertahanan siber organisasi banyak bertumpu pada tiga pendekatan konvensional: deteksi berbasis tanda tangan (signature-based detection), sistem berbasis aturan (rule-based system), dan deteksi anomali statistik klasik (anomaly-based detection). Ketiganya menjadi fondasi berbagai perangkat keamanan yang masih digunakan luas hingga saat ini, seperti intrusion detection system (IDS) berbasis signature (mis. Snort) maupun firewall dengan kebijakan akses statis (Frontiers in Computer Science, 2025; Network Intrusion Detection Study, 2025).

Signature-based detection bekerja dengan mencocokkan karakteristik trafik atau berkas terhadap basis data pola serangan yang telah diketahui sebelumnya. Pendekatan ini populer karena tingkat akurasinya yang tinggi dan risiko false positive yang rendah untuk ancaman yang telah dikenal. Sementara itu, rule-based system menerapkan seperangkat aturan logis yang ditetapkan administrator untuk mengizinkan atau memblokir aktivitas tertentu. Adapun anomaly-based detection berupaya mengatasi kekakuan

kedua pendekatan tersebut dengan membangun model "perilaku normal" sistem, lalu menandai penyimpangan sebagai indikasi ancaman (ResearchGate, 2025a; ResearchGate, 2025b).

Tabel 2.1 merangkum prinsip kerja, keunggulan, dan keterbatasan utama dari masing-masing pendekatan konvensional tersebut.

Tabel 2.1 Perbandingan Pendekatan Keamanan Siber Konvensional

Pendekatan	Prinsip Kerja	Keunggulan	Keterbatasan Utama
Signature-based detection	Mencocokkan pola trafik/berkas dengan basis data tanda tangan ancaman yang telah dikenal	Akurasi tinggi dan nyaris tanpa false positive untuk ancaman yang telah dikenal; ringan secara komputasi	Tidak mampu mendeteksi serangan zero-day atau varian baru yang belum memiliki signature (Frontiers in Computer Science, 2025; ResearchGate, 2025a)
Rule-based / firewall statis	Menerapkan aturan logis (izinkan/blokir) berdasarkan kriteria yang ditetapkan administrator	Transparan, mudah diaudit, dan performanya stabil secara real-time	Kaku terhadap pola serangan baru; memerlukan pembaruan manual berkelanjutan agar tetap relevan (Network Intrusion Detection Study, 2025)
Anomaly-based detection (statistik klasik)	Membangun model "perilaku normal" sistem, lalu menandai penyimpangan sebagai potensi ancaman	Berpotensi mendeteksi ancaman yang belum dikenal sebelumnya	Rawan menghasilkan false positive dalam jumlah besar; kesulitan memodelkan data dinamis berskala besar; tidak dapat menganalisis trafik terenkripsi (ResearchGate, 2025a; ResearchGate, 2025b)

2.2 Keterbatasan Pendekatan Berbasis Signature dan Rule

Keterbatasan paling mendasar dari deteksi berbasis signature adalah ketergantungannya pada basis data ancaman yang telah dikenal sebelumnya. Ketika penyerang menggunakan teknik atau varian malware baru yang belum tercatat dalam basis data, dikenal sebagai serangan **zero-day — sistem signature-based** pada dasarnya akan gagal mendeteksinya (Frontiers in Computer Science, 2025; ResearchGate, 2025a). Situasi ini diperparah oleh kebiasaan pelaku ancaman untuk membuat variasi kecil pada program serangan guna mengelabui signature yang sudah ada, sehingga menghasilkan varian baru yang secara fungsional identik namun tidak lagi cocok dengan pola lama.

Kajian terbaru terhadap evolusi network intrusion detection system (NIDS) juga menyoroti keterbatasan struktural lain: karena pencocokan signature bersifat parsial dan berbasis pola yang terisolasi, pendekatan ini kesulitan mengorelasikan rangkaian tahapan serangan yang membentuk satu kesatuan serangan siber (attack chain), misalnya dari pengintaian, intrusi, hingga pergerakan lateral (lateral movement) di dalam jaringan. Setiap tahapan tersebut, jika dilihat secara terpisah, dapat tampak sebagai aktivitas yang wajar sehingga lolos dari deteksi (Network Intrusion Detection Study, 2025).

Sementara itu, sistem rule-based menuntut administrator memperbarui aturan secara manual dan berkelanjutan agar tetap relevan terhadap pola serangan baru, sebuah proses yang tidak dapat mengimbangi kecepatan evolusi ancaman siber modern, terlebih di tengah percepatan otomatisasi serangan berbasis AI sebagaimana dibahas pada Bab 1.

2.3 Keterbatasan Deteksi Anomali Tradisional

Deteksi anomali berbasis statistik klasik dikembangkan untuk mengatasi kelemahan signature-based dalam menghadapi ancaman yang belum dikenal. Namun, pendekatan ini memiliki keterbatasannya sendiri. Pertama, membangun model "perilaku normal" yang akurat untuk lingkungan jaringan modern yang sangat dinamis dan berskala besar merupakan tantangan komputasi yang berat, sehingga model yang dihasilkan rentan menghasilkan alarm palsu (false positive) dalam jumlah besar (ResearchGate, 2025a; ResearchGate, 2025b).

Kedua, deteksi anomali statistik klasik umumnya kesulitan menganalisis trafik yang telah dienkripsi, sebuah keterbatasan signifikan mengingat mayoritas trafik internet saat ini menggunakan enkripsi TLS. Ketiga, volume alarm palsu yang tinggi berkontribusi pada fenomena "alert fatigue", yaitu kondisi ketika tim keamanan menjadi kelelahan dan cenderung mengabaikan sebagian notifikasi karena terlalu banyak peringatan yang ternyata bukan ancaman nyata, sehingga berisiko melewatkan insiden yang sesungguhnya kritis.

2.4 Keterbatasan Struktural: Fragmentasi Sistem dan Kesenjangan Talenta

Selain keterbatasan teknis pada masing-masing metode deteksi, terdapat pula keterbatasan struktural yang memperlemah efektivitas pertahanan siber konvensional secara keseluruhan. Pertama, banyak organisasi mengoperasikan berbagai perangkat dan platform keamanan yang tidak saling terintegrasi, menciptakan celah deteksi (visibility gap) yang memperlambat respons terhadap insiden dan memungkinkan serangan

menyebar lebih luas sebelum sempat ditangani (Constellation Research, 2025).

Kedua, industri keamanan siber global menghadapi kesenjangan keahlian yang semakin melebar. Studi tenaga kerja ISC2 tahun 2025 yang melibatkan lebih dari 16.000 profesional keamanan siber di seluruh dunia menemukan bahwa 59% organisasi melaporkan kesenjangan keahlian yang kritis atau signifikan, meningkat tajam dari 44% pada tahun sebelumnya. Keahlian terkait AI menempati posisi teratas kebutuhan keahlian yang paling mendesak (41%), diikuti oleh keamanan cloud (36%) (ISC2, 2025; Infosecurity Magazine, 2026). Studi yang sama menegaskan pergeseran penting: tantangan utama tim keamanan siber kini bukan lagi semata kekurangan jumlah personel, melainkan kekurangan keahlian spesifik untuk menghadapi ancaman yang semakin kompleks.

Kombinasi antara keterbatasan teknis metode deteksi konvensional dan keterbatasan struktural berupa fragmentasi sistem serta kesenjangan talenta inilah yang mendasari argumen mengapa pendekatan berbasis kecerdasan buatan menjadi kebutuhan, bukan sekadar pilihan, dalam keamanan siber kontemporer.

2.5 Mengapa Kecerdasan Buatan? Bukti Empiris dari Lapangan

Argumen mengenai pentingnya kecerdasan buatan dalam keamanan siber tidak hanya bersifat konseptual, tetapi juga didukung oleh data empiris berskala global. Laporan Cost of a Data Breach 2025 yang disusun IBM bersama Ponemon Institute, berdasarkan analisis mendalam terhadap insiden pelanggaran data nyata di berbagai organisasi menemukan bahwa organisasi yang menerapkan AI dan otomasi keamanan secara ekstensif

memperoleh manfaat signifikan dibandingkan organisasi yang tidak menerapkannya. Tabel 2.2 merangkum sejumlah temuan kunci.

Tabel 2.2 Dampak Penerapan AI dan Otomasi terhadap Siklus Pelanggaran Data (IBM Cost of a Data Breach 2025)

Indikator	Nilai
Rata-rata biaya pelanggaran data global (2025)	USD 4,44 juta (turun 9% dari USD 4,88 juta pada 2024)
Rata-rata siklus hidup pelanggaran data (identifikasi + penahanan)	241 hari — titik terendah dalam 9 tahun terakhir
Penghematan biaya bagi organisasi dengan penggunaan AI & otomasi keamanan yang ekstensif	±USD 1,9 juta per insiden
Percepatan siklus hidup pelanggaran pada organisasi pengguna AI & otomasi ekstensif	80 hari lebih cepat dibanding organisasi non-pengguna
Biaya pelanggaran: terdeteksi < 200 hari vs > 200 hari	USD 3,61 juta vs USD 5,49 juta (selisih ±USD 1,88 juta)

Temuan ini memberikan justifikasi kuantitatif yang kuat: kecerdasan buatan bukan sekadar tren teknologi, melainkan faktor yang secara terukur mempercepat deteksi, mempersingkat waktu penahanan insiden, dan menekan kerugian finansial akibat pelanggaran data. Efek ini konsisten dengan temuan pada laporan-laporan tahun sebelumnya, yang secara berulang menunjukkan bahwa organisasi dengan penerapan AI dan otomasi keamanan yang matang secara signifikan lebih unggul dalam kecepatan respons dibandingkan organisasi yang masih mengandalkan proses manual (IBM Security & Ponemon Institute, 2025; Help Net Security, 2025).

Di sisi lain, kecerdasan buatan juga menawarkan jawaban terhadap keterbatasan structural yang dibahas pada subbab 2.4: AI dan otomasi memungkinkan tim keamanan siber yang terbatas jumlahnya untuk tetap memantau volume trafik dan potensi ancaman yang jauh melampaui kapasitas analisis manual, sekaligus mengurangi beban kerja repetitif yang berkontribusi pada kelelahan analis (Infosecurity Magazine, 2026).

Catatan: Mengapa Kecerdasan Buatan Penting dalam Keamanan Siber

- *Organisasi yang menerapkan AI dan otomasi keamanan secara ekstensif mencatat penghematan biaya pelanggaran data rata-rata USD 1,9 juta per insiden serta siklus deteksi-penahanan yang lebih singkat 80 hari dibandingkan organisasi yang tidak menerapkannya (IBM Security & Ponemon Institute, 2025; Help Net Security, 2025).*
- *Rata-rata biaya pelanggaran data global justru menurun untuk pertama kalinya dalam lima tahun pada 2025, dan penurunan ini sebagian besar didorong oleh deteksi dan penahanan insiden yang lebih cepat berkat AI dan otomasi keamanan (Help Net Security, 2025).*
- *Kesenjangan talenta keamanan siber semakin bergeser dari sekadar kekurangan jumlah personel menjadi kekurangan keahlian spesifik, 59% organisasi melaporkan kesenjangan keahlian yang kritis atau signifikan pada 2025, dengan keahlian AI menjadi kebutuhan paling mendesak (ISC2, 2025; Infosecurity Magazine, 2026).*
- *AI generatif dan model bahasa besar kini mulai dimanfaatkan sebagai kolaborator dalam sistem deteksi intrusi jaringan, melengkapi keterbatasan pendekatan signature-based dan rule-based dalam mengenali pola serangan bertahap (multi-stage attack) (Network Intrusion Detection Study, 2025).*

- *Kombinasi explainable AI (XAI) dengan model deteksi berbasis machine learning mulai diadopsi secara luas untuk mengatasi persoalan "black box", sehingga rekomendasi keamanan yang dihasilkan AI dapat dipahami dan dipercaya oleh analis keamanan (Frontiers in Computer Science, 2025).*

Catatan: AI Bukan Solusi Tanpa Risiko

- *Laporan IBM 2025 juga menemukan bahwa 97% organisasi yang mengalami insiden terkait AI tidak memiliki kontrol akses AI yang memadai, menunjukkan bahwa adopsi AI tanpa tata kelola justru dapat menciptakan "utang keamanan" (security debt) baru (IBM Security & Ponemon Institute, 2025; Help Net Security, 2025).*
- *Model AI, khususnya deep learning, umumnya bersifat black box sehingga keputusannya sulit dijelaskan, sebuah tantangan yang mendorong berkembangnya subbidang explainable AI dalam keamanan siber (Frontiers in Computer Science, 2025).*
- *AI generatif yang sama juga dimanfaatkan pelaku ancaman untuk menghasilkan malware polimorfik dan phishing yang lebih meyakinkan, sehingga adopsi AI oleh tim keamanan harus diimbangi kewaspadaan terhadap penyalahgunaan AI oleh pihak yang tidak bertanggung jawab (dibahas lebih lanjut pada Bab 1 dan Bab 12).*

2.6 Menuju Pendekatan Adaptif: Dari Deteksi Statis ke Pembelajaran Kontekstual

Pergeseran dari pendekatan konvensional menuju pendekatan berbasis kecerdasan buatan pada dasarnya adalah pergeseran paradigma: dari sistem yang "mengetahui" ancaman berdasarkan pola yang telah didefinisikan sebelumnya, menuju sistem yang "belajar" mengenali pola dan konteks ancaman secara berkelanjutan dari data. Kajian bibliometrik terbaru terhadap publikasi AI-driven cybersecurity periode 2015–2025 mengidentifikasi empat domain utama penerapan AI yang berkembang paling pesat: deteksi intrusi berbasis anomali, analisis malware otomatis, pencegahan phishing dan rekayasa sosial, serta security orchestration, automation and response (SOAR), dengan model machine learning dan deep learning secara konsisten menunjukkan kinerja yang lebih unggul dibandingkan pendekatan rule-based tradisional dalam mendeteksi ancaman yang kompleks dan terus berevolusi (ResearchGate, 2025b).

Perkembangan lebih lanjut bahkan mulai mengarah pada kolaborasi antara model neural network dengan large language model (LLM) untuk memperkuat kemampuan NIDS dalam memahami konteks serangan secara lebih menyeluruh (Network Intrusion Detection Study, 2025). Fondasi teknis dari pergeseran paradigma inilah mencakup machine learning, deep learning, hingga pendekatan berbasis graf seperti Graph Neural Networks yang akan dibahas secara mendalam pada Bagian II buku ini.

2.7 Rangkuman

Pendekatan keamanan siber konvensional signature-based, rule-based, dan anomaly-based statistik klasik memiliki keterbatasan mendasar dalam menghadapi ancaman zero-day, serangan bertahap (multi-stage), dan trafik terenkripsi. Keterbatasan teknis tersebut diperparah oleh keterbatasan struktural berupa fragmentasi sistem keamanan dan kesenjangan keahlian tenaga kerja, yang menurut studi ISC2 2025 kini menjadi tantangan yang

lebih mendesak dibandingkan sekadar kekurangan jumlah personel. Data empiris dari laporan IBM Cost of a Data Breach 2025 menunjukkan bahwa organisasi yang menerapkan AI dan otomasi keamanan secara ekstensif memperoleh penghematan biaya insiden rata-rata USD 1,9 juta dan siklus deteksi yang lebih cepat 80 hari. Kecerdasan buatan menawarkan pergeseran paradigma dari deteksi statis berbasis pola yang telah dikenal menuju pembelajaran kontekstual yang adaptif terhadap ancaman baru. Adopsi AI dalam keamanan siber tetap perlu diimbangi tata kelola yang memadai, mengingat risiko "black box", kurangnya kontrol akses AI, dan potensi penyalahgunaan AI generatif oleh pelaku ancaman itu sendiri.

Daftar Pustaka

Frontiers in Computer Science. (2025). Evaluating Machine Learning-Based Intrusion Detection Systems with Explainable AI: Enhancing Transparency and Interpretability. Frontiers in Computer Science, 7.

Network Intrusion Detection: Evolution from Conventional Approaches to LLM Collaboration and Emerging Risks. (2025). arXiv:2510.23313.

ResearchGate. (2025a). A Comparative Analysis of Signature-Based and Anomaly-Based Intrusion Detection Systems.

Studi bibliometrik AI-driven cybersecurity 2015–2025 (systematic literature review). (2025b). Dikutip dalam ResearchGate, Juni 2025.

IBM Security & Ponemon Institute. (2025). Cost of a Data Breach Report 2025.

Help Net Security. (2025). Average Global Data Breach Cost Now $4.44 Million, 4 Agustus 2025.

ISC2. (2025). 2025 Cybersecurity Workforce Study.

Infosecurity Magazine. (2026). Skills Shortages Trump Headcount as Critical Cyber Challenge, dikutip 16 Februari 2026.

Constellation Research. (2025). Analisis fragmentasi sistem keamanan dan pergeseran menuju zero trust, dikutip dalam Beritajejakfakta.id, 14 April 2026.

BAB 3

Taksonomi Kecerdasan Buatan dalam Keamanan Siber

3.1 Perlunya Peta Taksonomi

Bab 1 dan Bab 2 telah menunjukkan bahwa lanskap ancaman siber modern menuntut pendekatan pertahanan yang adaptif, dan bahwa kecerdasan buatan menawarkan jawaban yang didukung bukti empiris. Namun, istilah "kecerdasan buatan" sering digunakan secara longgar untuk merujuk pada beragam teknik yang sebenarnya berbeda secara fundamental — mulai dari algoritma statistik klasik hingga model bahasa berskala besar. Sebuah kajian bibliometrik terhadap lebih dari 81.000 artikel ilmiah pada basis data Scopus menunjukkan betapa luas dan beragamnya penerapan machine learning dalam keamanan siber, sehingga diperlukan kerangka klasifikasi yang jelas untuk memahami posisi dan keterkaitan masing-masing pendekatan (ACM Computing Surveys, 2025).

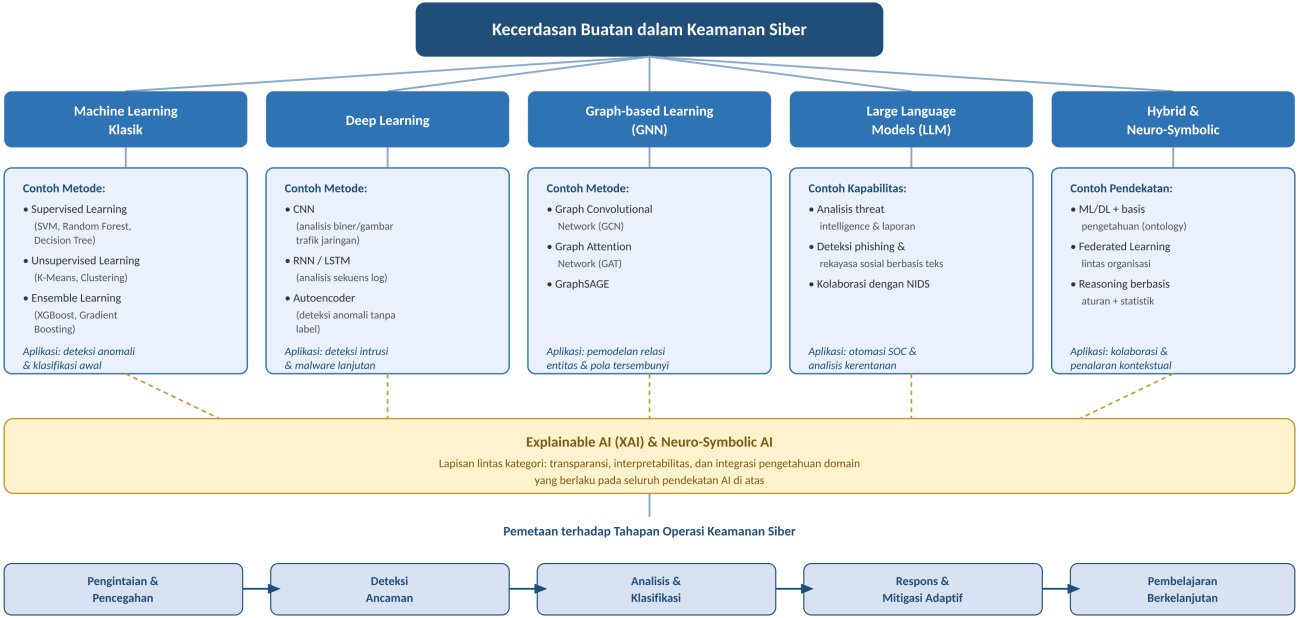
Bab ini menyajikan peta taksonomi kecerdasan buatan dalam konteks keamanan siber, disusun dalam lima kategori utama: machine learning klasik, deep learning, graph-based learning, large language model (LLM), serta pendekatan hybrid dan neuro-symbolic. Kelima kategori tersebut dilengkapi lapisan lintas kategori berupa explainable AI (XAI), yang berperan penting dalam menjembatani kinerja model dengan kebutuhan

interpretabilitas oleh analis keamanan (ACM Computing Surveys, 2025; arXiv, 2025).

3.2 Peta Taksonomi: Gambaran Menyeluruh

Gambar 3.1 menyajikan peta taksonomi kecerdasan buatan dalam keamanan siber secara visual, mencakup kategori utama, contoh metode representatif pada masing-masing kategori, lapisan XAI dan neuro-symbolic yang bersifat lintas kategori, serta pemetaannya terhadap tahapan operasi keamanan siber.

Gambar 3.1 Taksonomi Kecerdasan Buatan dalam Keamanan Siber



Sumber: disintesis oleh penulis dari ACM Computing Surveys (2025); Discover Artificial Intelligence, Springer (2025); ScienceDirect Systematic Review on LLM in Cybersecurity (2026); arXiv Neuro-Symbolic AI for Cybersecurity (2025).

Gambar 3.1 Taksonomi Kecerdasan Buatan dalam Keamanan Siber (disintesis oleh penulis)

Catatan: Membaca Peta Taksonomi Bab Ini

- *Taksonomi pada Gambar 3.1 disusun berdasarkan sintesis dari kajian literatur (survei ACM Computing Surveys 2025, Discover Artificial Intelligence/Springer 2025, systematic review LLM in Cybersecurity 2026, dan survei Neuro-Symbolic AI for Cybersecurity 2025) dan bersifat fungsional, bukan eksklusif dalam praktiknya, sistem keamanan siber modern kerap menggabungkan lebih dari satu kategori sekaligus (ACM Computing Surveys, 2025; Discover Artificial Intelligence, 2025; ScienceDirect, 2026; Yan et al., 2023).*
- *Explainable AI (XAI) dan pendekatan neuro-symbolic ditempatkan sebagai lapisan lintas kategori karena keduanya tidak berdiri sebagai metode deteksi yang berdiri sendiri, melainkan berfungsi memperkuat transparansi dan interpretabilitas seluruh kategori AI di atasnya (ACM Computing Surveys, 2025; arXiv, 2025).*
- *Baris paling bawah pada Gambar 3.1 memetakan taksonomi terhadap siklus operasi keamanan siber pencegahan, deteksi, analisis, respons, hingga pembelajaran berkelanjutan untuk menunjukkan bagaimana setiap kategori AI berkontribusi pada tahapan yang berbeda dalam pertahanan siber.*

3.3 Machine Learning Klasik

Machine learning klasik merupakan fondasi paling awal penerapan kecerdasan buatan dalam keamanan siber. Kategori ini mencakup algoritma seperti support vector machine (SVM), random forest, decision tree untuk pembelajaran terawasi (supervised learning), serta k-means dan berbagai teknik clustering untuk pembelajaran tak terawasi (unsupervised learning). Karakteristik utamanya adalah kebutuhan akan rekayasa fitur

(feature engineering) secara manual pakar keamanan perlu menentukan sendiri atribut data mana yang relevan sebagai masukan model, namun dengan kebutuhan komputasi yang relatif ringan dibandingkan pendekatan yang lebih kompleks.

Pendekatan ini tetap relevan hingga saat ini, khususnya untuk kasus dengan data terstruktur berskala menengah dan kebutuhan interpretabilitas yang tinggi, seperti pada sistem deteksi anomali statistik dan klasifikasi malware tahap awal.

3.4 Deep Learning

Berbeda dengan machine learning klasik, deep learning mampu mempelajari representasi fitur secara otomatis langsung dari data mentah melalui lapisan-lapisan jaringan saraf tiruan. Beberapa arsitektur yang banyak diterapkan dalam keamanan siber meliputi convolutional neural network (CNN) untuk analisis representasi biner atau citra trafik jaringan, recurrent neural network (RNN) dan long short-term memory (LSTM) untuk menganalisis data sekuensial seperti log sistem dan aliran trafik dari waktu ke waktu, serta autoencoder untuk deteksi anomali tanpa memerlukan data berlabel.

Kajian komprehensif terhadap intrusion detection system menunjukkan bahwa integrasi machine learning dan deep learning secara signifikan meningkatkan kemampuan IDS dalam mendeteksi ancaman siber yang terus berkembang, sekaligus menghadirkan tantangan baru terkait kebutuhan data pelatihan berskala besar dan interpretabilitas model (Discover Artificial Intelligence, 2025).

3.5 Graph-based Learning: Graph Neural Networks (GNN)

Kategori ini merepresentasikan pergeseran cara pandang terhadap data keamanan siber: alih-alih memperlakukan setiap entitas (host, pengguna, paket data) sebagai baris data yang independen, graph-based learning memodelkan seluruh entitas dan relasinya sebagai satu struktur graf yang saling terhubung. Pendekatan ini memungkinkan model seperti graph convolutional network (GCN), graph attention network (GAT), dan GraphSAGE untuk menangkap pola relasional yang kompleks, misalnya pergerakan lateral penyerang dari satu host ke host lain, yang sulit dideteksi oleh model berbasis tabel data konvensional.

Sebuah survei tentang graph mining untuk keamanan siber menegaskan bahwa kemampuan graph-based learning dalam memodelkan relasi antarentitas menjadikannya pendekatan yang sangat relevan untuk mendeteksi ancaman siber yang bersifat multi-tahap dan tersembunyi dalam struktur jaringan yang kompleks (Yan et al., 2023). Pendekatan inilah yang menjadi fokus studi kasus implementasi pada Bagian III buku ini.

3.6 Large Language Model (LLM) dalam Keamanan Siber

Perkembangan model bahasa besar (LLM) berbasis arsitektur transformer membuka kapabilitas baru dalam keamanan siber, khususnya dalam memproses data tidak terstruktur berbasis bahasa alami dan kode pemrograman. Sebuah tinjauan sistematis terhadap 167 studi yang diterbitkan antara 2022–2025 mengidentifikasi bahwa LLM kini menjadi komponen integral ekosistem keamanan siber modern, memperkuat kapabilitas pertahanan sekaligus memunculkan kelas ancaman baru berbasis konten yang dihasilkan AI (AI-Generated Content) (ScienceDirect, 2026).

Beberapa kapabilitas LLM yang relevan meliputi analisis dan peringkasan laporan threat intelligence, deteksi phishing dan rekayasa sosial berbasis analisis teks, otomasi triase pada security operations center (SOC), hingga kolaborasi dengan model neural network lain, misalnya kombinasi model GPT dan BERT dengan classifier berbasis LSTM, untuk memperkuat kemampuan network intrusion detection system dalam memahami konteks serangan secara lebih menyeluruh, sebagaimana telah disinggung pada Bab 2 (ScienceDirect, 2026).

3.7 Pendekatan Hybrid dan Neuro-Symbolic

Kategori kelima mencakup pendekatan yang menggabungkan kekuatan pembelajaran statistik (machine learning/deep learning) dengan basis pengetahuan eksplisit berupa aturan, ontologi, atau kerangka penalaran simbolik. Survei terbaru mengenai neuro-symbolic AI untuk keamanan siber menunjukkan bahwa pendekatan hybrid semacam ini berpotensi mengatasi keterbatasan murni statistik seperti kesulitan menjelaskan keputusan model sekaligus mempertahankan kemampuan beradaptasi terhadap pola data baru (arXiv, 2025).

Termasuk dalam kategori ini adalah federated learning, yaitu skema pembelajaran terdistribusi yang memungkinkan beberapa organisasi melatih model bersama tanpa perlu saling berbagi data mentah, pendekatan yang sangat relevan untuk kolaborasi keamanan siber lintas-organisasi mengingat sensitivitas data ancaman yang dimiliki masing-masing pihak.

3.8 Rangkuman Taksonomi dalam Tabel

Tabel 3.1 merangkum karakteristik utama dan contoh kasus penggunaan dari kelima kategori kecerdasan buatan yang telah dibahas.

Tabel 3.1 Ringkasan Taksonomi Kecerdasan Buatan dalam Keamanan Siber

Kategori	Karakteristik Utama	Contoh Kasus Penggunaan
Machine Learning Klasik	Membutuhkan rekayasa fitur (feature engineering) manual; efisien secara komputasi; cocok untuk data terstruktur berskala menengah	Deteksi anomali statistik, klasifikasi awal malware, penyaringan spam
Deep Learning	Belajar representasi fitur secara otomatis dari data mentah berskala besar; menuntut sumber daya komputasi lebih tinggi	Deteksi intrusi lanjutan, klasifikasi malware kompleks, analisis citra biner
Graph-based Learning (GNN)	Memodelkan data sebagai struktur graf (node & edge) sehingga mampu menangkap relasi antarentitas jaringan	Deteksi pola serangan tersembunyi, analisis lateral movement, rekomendasi mitigasi kontekstual
Large Language Model (LLM)	Memproses dan menghasilkan teks berskala besar; mampu memahami konteks bahasa alami dan kode	Analisis threat intelligence, deteksi phishing berbasis teks, otomasi SOC
Hybrid & Neuro-Symbolic	Menggabungkan pembelajaran statistik dengan basis pengetahuan/aturan eksplisit	Federated learning lintas organisasi, penalaran kontekstual berbasis kebijakan keamanan

3.9 Implikasi Taksonomi bagi Struktur Buku

Peta taksonomi pada bab ini menjadi kerangka acuan bagi struktur pembahasan pada Bagian II buku ini. Bab 4 akan membahas secara khusus penerapan machine learning untuk deteksi anomali dan intrusi; Bab 5 mengulas deep learning dalam analisis malware dan traffic jaringan; Bab 6 membahas graph neural networks secara mendalam sebagai fondasi bagi studi kasus pada Bagian III; dan Bab 7 mengulas explainable AI serta large language model dalam konteks keamanan siber. Dengan kerangka taksonomi ini, pembaca diharapkan dapat memahami posisi setiap teknik AI secara proporsional, alih-alih memperlakukan "AI untuk keamanan siber" sebagai satu entitas tunggal yang homogen.

3.10 Rangkuman

Kecerdasan buatan dalam keamanan siber bukan satu entitas tunggal, melainkan mencakup beragam pendekatan dengan karakteristik dan kasus penggunaan yang berbeda-beda. Lima kategori utama taksonomi meliputi machine learning klasik, deep learning, graph-based learning (GNN), large language model, serta pendekatan hybrid dan neuro-symbolic. Explainable AI (XAI) dan pendekatan neuro-symbolic berperan sebagai lapisan lintas kategori yang memperkuat transparansi dan interpretabilitas seluruh pendekatan AI di atasnya. Setiap kategori memiliki relevansi pada tahapan operasi keamanan siber yang berbeda, mulai dari pencegahan, deteksi, analisis, respons, hingga pembelajaran berkelanjutan. Graph-based learning, khususnya Graph Neural Networks, menjadi fokus khusus pembahasan pada bagian selanjutnya buku ini karena relevansinya dengan studi kasus implementasi yang diangkat penulis.

Daftar Pustaka

ACM Computing Surveys. (2025). Machine Learning for Cybersecurity: A Comprehensive Literature Review.

Discover Artificial Intelligence, Springer Nature. (2025). A Comprehensive Survey on Intrusion Detection Systems with Advances in Machine Learning, Deep Learning and Emerging Cybersecurity Challenges.

ScienceDirect. (2026). Large Language Models for Cybersecurity Intelligence: A Systematic Review of Emerging Threats, Defensive Capabilities, and Security Evaluation Frameworks.

Yan, B., Yang, C., Shi, C., Fang, Y., Li, Q., Ye, Y., & Du, J. (2023). Graph Mining for Cybersecurity: A Survey. ACM Transactions on Knowledge Discovery from Data (TKDD), 18(2), 1–52. Dikutip dalam arXiv:2512.12112 (2025).

arXiv. (2025). Neuro-Symbolic AI for Cybersecurity: State of the Art, Challenges, and Opportunities. arXiv:2509.06921.

BAB 4

Machine Learning untuk Deteksi Anomali dan Intrusi

4.1 Pendahuluan

Sebagaimana dipetakan pada taksonomi di Bab 3, machine learning klasik menjadi fondasi paling matang dan paling banyak diadopsi dalam penerapan kecerdasan buatan untuk keamanan siber. Bab ini membahas secara khusus bagaimana algoritma machine learning diterapkan untuk dua tugas inti: deteksi anomali (anomaly detection) dan deteksi intrusi (intrusion detection). Pembahasan mencakup prinsip kerja, algoritma populer, dataset benchmark yang umum digunakan peneliti, serta dua studi kasus konkret yang menggambarkan penerapan nyata pendekatan ini.

4.2 Landasan Fundamental Matematis

Sebelum membahas penerapan praktis, penting untuk memahami fondasi matematis yang mendasari algoritma machine learning yang digunakan dalam deteksi anomali dan intrusi. Pemahaman ini membantu pembaca tidak sekadar menggunakan algoritma sebagai kotak hitam, melainkan memahami mengapa suatu algoritma bekerja dan kapan algoritma tersebut paling sesuai digunakan.

4.2.1 Formulasi Umum Masalah Klasifikasi

Secara formal, masalah deteksi intrusi berbasis supervised learning dapat dirumuskan sebagai pencarian fungsi f yang memetakan ruang fitur X ke ruang label Y, berdasarkan himpunan data pelatihan berlabel. Diberikan himpunan data D = {(x_1,y_1), (x_2,y_2), …, (x_n,y_n)}, di mana setiap x_i merupakan vektor fitur berdimensi d yang merepresentasikan atribut trafik jaringan (misalnya durasi koneksi, jumlah byte terkirim, protokol yang digunakan), dan y_i ∈ {normal, serangan} merupakan label kelas, tujuan pembelajaran adalah menemukan fungsi f: X → Y yang meminimalkan risiko empiris berikut:

$$R(f) = (1/n)\,\Sigma_{i=1}{}^{n}\, L(f(x_i), y_i)$$

dengan L merupakan fungsi kerugian (loss function) yang mengukur selisih antara prediksi model $f(x_i)$ dan label sebenarnya y_i. Berbagai algoritma yang dibahas pada bab ini pada dasarnya menawarkan pendekatan berbeda untuk mendekati fungsi f secara optimal.

4.2.2 Support Vector Machine (SVM)

SVM bekerja dengan mencari hyperplane pemisah yang memaksimalkan margin antara dua kelas data. Untuk kasus klasifikasi biner yang dapat dipisahkan secara linear, hyperplane didefinisikan sebagai:

$$w \cdot x + b = 0$$

dengan w adalah vektor bobot yang tegak lurus terhadap hyperplane, dan b adalah bias. SVM mencari w dan b yang memaksimalkan margin $2/\|w\|$ dengan kendala bahwa seluruh titik data diklasifikasikan dengan benar, dirumuskan sebagai masalah optimasi:

$$min\ (1/2)\|w\|^2 \text{ dengan kendala } y_i(w \cdot x_i + b) \geq 1 \text{ untuk semua } i$$

Untuk data yang tidak dapat dipisahkan secara linear — kondisi yang umum terjadi pada data trafik jaringan nyata, SVM menggunakan kernel trick,

memetakan data ke ruang berdimensi lebih tinggi melalui fungsi kernel $\phi(x)$, tanpa perlu menghitung transformasi tersebut secara eksplisit. Kernel yang umum digunakan pada deteksi intrusi antara lain kernel Radial Basis Function (RBF): $K(x_i,x_j) = \exp(-\gamma\|x_i-x_j\|^2)$, yang memungkinkan SVM menangkap batas keputusan (decision boundary) yang non-linear dan kompleks.

4.2.3 Decision Tree dan Random Forest

Decision tree membangun struktur pohon keputusan dengan memilih fitur pemisah pada setiap node yang paling efektif memisahkan kelas data. Efektivitas pemisahan diukur menggunakan entropy atau Gini index. Entropy suatu himpunan data S didefinisikan sebagai:

$$Entropy(S) = -\Sigma_i\, p_i\, log_2(p_i)$$

dengan p_i adalah proporsi data pada kelas i dalam himpunan S. Algoritma memilih fitur pemisah yang memaksimalkan information gain, yaitu penurunan entropy setelah pemisahan dilakukan:

$$IG(S, A) = Entropy(S) - \Sigma_v\, (|S_v|/|S|) \cdot Entropy(S_v)$$

dengan A adalah fitur yang dievaluasi dan S_v adalah subset data hasil pemisahan berdasarkan nilai v dari fitur A. Random Forest memperkuat pendekatan ini melalui teknik bagging (bootstrap aggregating): membangun banyak decision tree secara independen, masing-masing dilatih pada subset data hasil pengambilan sampel acak dengan pengembalian (bootstrap sample), serta hanya mempertimbangkan subset fitur acak pada setiap pemisahan. Prediksi akhir ditentukan melalui voting mayoritas dari seluruh pohon, yang secara signifikan mengurangi risiko overfitting dibandingkan satu decision tree tunggal.

4.2.4 Gradient Boosting (XGBoost)

Berbeda dengan Random Forest yang membangun pohon secara independen dan paralel, gradient boosting membangun model secara sekuensial, dengan setiap pohon baru dilatih untuk memperbaiki kesalahan (residual) dari kombinasi pohon-pohon sebelumnya. Secara formal, model prediksi pada iterasi ke-t dirumuskan sebagai:

$$F_t(x) = F_{t-1}(x) + \eta \cdot h_t(x)$$

dengan $F_{t-1}(x)$ adalah prediksi gabungan dari model sebelumnya, $h_t(x)$ adalah decision tree baru yang dilatih untuk memprediksi gradien negatif dari fungsi kerugian terhadap prediksi saat ini, dan η adalah learning rate yang mengontrol kontribusi setiap pohon baru. XGBoost menyempurnakan pendekatan ini dengan menambahkan regularisasi eksplisit untuk mencegah overfitting serta optimasi komputasi yang memungkinkan pelatihan pada dataset berskala besar secara efisien, dua faktor yang menjelaskan performanya yang konsisten unggul pada studi kasus di Subbab 4.6.

4.2.5 K-Means Clustering

Sebagai representasi pembelajaran tak terawasi, K-Means bertujuan mempartisi n titik data ke dalam k kelompok (cluster) sedemikian rupa sehingga jumlah kuadrat jarak setiap titik terhadap centroid kelompoknya diminimalkan:

$$J = \Sigma_{k=1}{}^{K} \Sigma_{x \in C_k} \|x - \mu_k\|^2$$

dengan μ_k adalah centroid (titik pusat) kelompok C_k. Algoritma bekerja secara iteratif: (1) menetapkan setiap titik data ke centroid terdekat, dan (2) memperbarui posisi centroid sebagai rata-rata seluruh titik dalam kelompoknya, hingga konvergen. Dalam konteks deteksi anomali, titik data yang berada jauh dari seluruh centroid (memiliki jarak yang secara

signifikan lebih besar dari rata-rata) dapat diinterpretasikan sebagai potensi anomali.

4.2.6 k-Nearest Neighbors (k-NN)

k-NN mengklasifikasikan sebuah titik data baru berdasarkan mayoritas kelas dari k titik data pelatihan yang paling dekat dengannya, biasanya diukur menggunakan jarak Euclidean:

$$d(x, x_i) = \sqrt{(\Sigma_{j=1}^{d} (x_j - x_{ij})^2)}$$

Kesederhanaan k-NN menjadikannya dasar konseptual yang berguna untuk memahami deteksi anomali berbasis kedekatan (proximity-based), meskipun pada dataset berskala besar seperti trafik jaringan produksi, k-NN memerlukan optimasi struktur data (seperti k-d tree) untuk tetap efisien secara komputasi.

4.2.7 Metrik Evaluasi

Evaluasi kinerja model deteksi intrusi menggunakan sejumlah metrik standar yang dihitung dari matriks konfusi (confusion matrix), yaitu tabel yang membandingkan prediksi model dengan label sebenarnya melalui empat kategori: true positive (TP), true negative (TN), false positive (FP), dan false negative (FN).

$$Akurasi = (TP + TN) / (TP + TN + FP + FN)$$

$$Presisi = TP / (TP + FP) \qquad Recall = TP / (TP + FN)$$

$$F1\text{-}Score = 2 \times (Presisi \times Recall) / (Presisi + Recall)$$

Presisi mengukur proporsi prediksi "serangan" yang benar-benar merupakan serangan (relevan untuk meminimalkan false alarm), sementara recall mengukur proporsi serangan aktual yang berhasil terdeteksi (relevan untuk meminimalkan serangan yang lolos deteksi). F1-

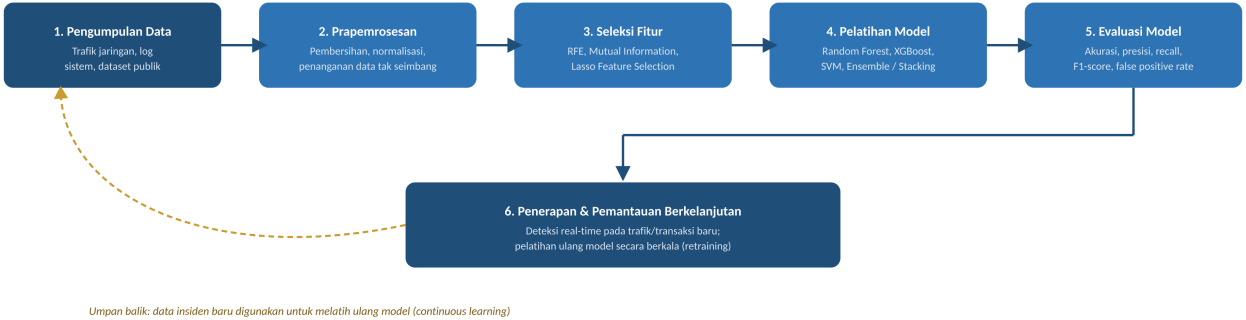
score menyeimbangkan keduanya, dan menjadi metrik yang paling sering dirujuk pada studi kasus di bab ini karena data keamanan siber pada praktiknya sangat tidak seimbang (jumlah trafik normal jauh melampaui jumlah serangan).

4.3 Prinsip Kerja Deteksi Berbasis Machine Learning

Secara umum, deteksi intrusi berbasis machine learning dapat dikategorikan menjadi dua pendekatan besar. Pertama, pembelajaran terawasi (supervised learning), yang melatih model menggunakan data berlabel, setiap sampel trafik ditandai sebagai "normal" atau jenis serangan tertentu, sehingga model belajar memetakan pola fitur terhadap label yang sesuai. Kedua, pembelajaran tak terawasi (unsupervised learning), yang tidak memerlukan data berlabel dan alih-alih berupaya mengenali struktur atau kelompok tersembunyi dalam data, sehingga cocok untuk mendeteksi pola anomali yang belum pernah didefinisikan sebelumnya.

Terlepas dari pendekatan yang dipilih, pembangunan sistem deteksi intrusi berbasis machine learning secara umum mengikuti alur kerja yang sistematis, sebagaimana diilustrasikan pada Gambar 4.1.

Gambar 4.1 Alur Kerja Sistem Deteksi Intrusi Berbasis Machine Learning



Sumber: disintesis oleh penulis dari ScienceDirect (2025), Nature Scientific Reports (2025), dan PMC (2025).

Gambar 4.1 Alur Kerja Sistem Deteksi Intrusi Berbasis Machine Learning (disintesis oleh penulis)

Alur kerja ini dimulai dari pengumpulan data mentah (trafik jaringan, log sistem, atau dataset publik), dilanjutkan dengan prapemrosesan untuk membersihkan dan menormalisasi data, seleksi fitur untuk memilih atribut paling relevan, pelatihan model, evaluasi kinerja menggunakan metrik seperti akurasi, presisi, recall, dan F1-score, hingga penerapan model pada lingkungan produksi disertai pemantauan dan pelatihan ulang berkelanjutan seiring munculnya pola ancaman baru.

4.4 Algoritma Machine Learning Populer untuk IDS

Tabel 4.1 merangkum sejumlah algoritma machine learning yang paling umum digunakan dalam penelitian dan penerapan sistem deteksi intrusi, beserta karakteristik dan contoh penerapannya.

Tabel 4.1 Algoritma Machine Learning Populer untuk Deteksi Intrusi

Algoritma	Jenis Pembelajaran	Karakteristik	Contoh Penerapan pada IDS
Support Vector Machine (SVM)	Supervised	Efektif pada data berdimensi tinggi; sensitif terhadap pemilihan kernel	Klasifikasi biner trafik normal vs. serangan
Decision Tree	Supervised	Mudah diinterpretasi; rawan overfitting pada data kompleks	Deteksi cepat dengan aturan keputusan eksplisit
Random Forest	Supervised (ensemble)	Menggabungkan banyak decision tree; tahan terhadap overfitting dan noise	Deteksi intrusi multi-kelas dengan akurasi tinggi

XGBoost	Supervised (boosting)	Membangun model secara bertahap untuk memperbaiki kesalahan sebelumnya; performa tinggi pada data tabular	Deteksi anomali pada dataset IoT dan jaringan besar
K-Means Clustering	Unsupervised	Mengelompokkan data tanpa label berdasarkan kemiripan fitur	Segmentasi pola trafik untuk identifikasi anomali awal
Autoencoder	Unsupervised	Belajar merekonstruksi data normal; anomali dikenali dari error rekonstruksi tinggi	Deteksi zero-day pada data yang belum pernah terlihat sebelumnya

4.5 Dataset Benchmark untuk Penelitian IDS

Penelitian akademik di bidang deteksi intrusi umumnya menggunakan sejumlah dataset benchmark yang telah tersedia secara publik agar hasil penelitian dapat dibandingkan secara objektif antarstudi. Tabel 4.2 merangkum dataset yang paling sering digunakan pada literatur terbaru.

Tabel 4.2 Dataset Benchmark yang Umum Digunakan pada Penelitian IDS

Dataset	Karakteristik	Contoh Kelas Serangan
NSL-KDD	Versi perbaikan dari KDD Cup 1999; 41 fitur; standar akademik yang banyak digunakan	DoS, Probe, User-to-Root (U2R), Remote-to-Local (R2L)
CICIDS2017 / CSE-CIC-IDS2018	Trafik jaringan realistis yang dihasilkan Canadian Institute for Cybersecurity; mencakup skenario serangan modern	Brute Force, Infiltration, Web Attack, DDoS, Botnet

UNSW-NB15	Kombinasi trafik normal dan sintetik dengan pola serangan kontemporer	Fuzzers, Backdoor, Exploits, Generic, Shellcode
CIC-MalMem-2022	Data memory dump untuk deteksi malware tersembunyi (fileless malware)	Ransomware, Trojan, Spyware berbasis analisis memori

4.6 Studi Kasus 1: Deteksi Intrusi Jaringan pada Dataset NSL-KDD

Konteks dan Permasalahan

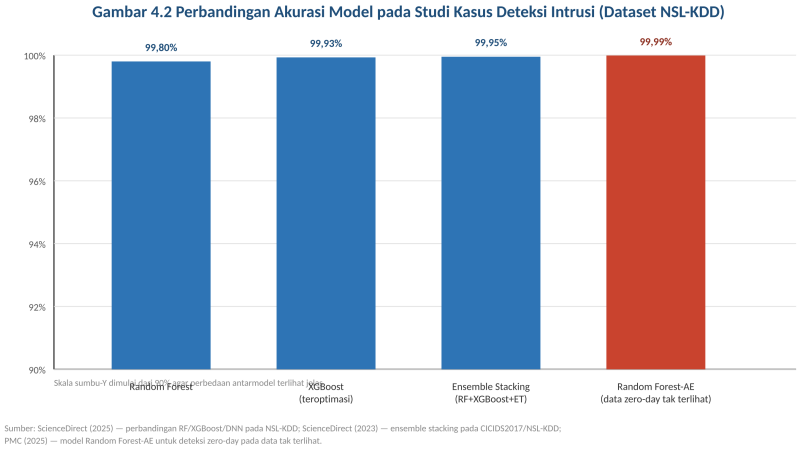
Sebagai ilustrasi penerapan konsep pada subbab sebelumnya, bagian ini menyajikan sintesis dari beberapa studi terbaru yang menerapkan machine learning pada dataset NSL-KDD, salah satu dataset benchmark paling banyak digunakan untuk penelitian IDS. Permasalahan yang diangkat adalah bagaimana mengklasifikasikan trafik jaringan ke dalam kategori normal atau salah satu jenis serangan (DoS, Probe, U2R, R2L) dengan akurasi setinggi mungkin, sekaligus menjaga model tetap mampu mengenali varian serangan yang belum pernah terlihat sebelumnya (zero-day).

Pendekatan

Studi pertama menerapkan model Random Forest tunggal yang dioptimasi, mencapai akurasi 99,80% dengan nilai AUC 0,9988 pada dataset NSL-KDD (ScienceDirect, 2025b). Studi kedua mengoptimasi model XGBoost menggunakan penyetelan hyperparameter, mencapai akurasi 99,93% dengan F1-score 99,84% dan tingkat false positive yang sangat rendah (Nature Scientific Reports, 2025). Studi ketiga menerapkan pendekatan

ensemble stacking yang menggabungkan Random Forest, XGBoost, dan Extra-Trees dengan meta-model regresi logistik, dipadukan dengan teknik seleksi fitur Recursive Feature Elimination (RFE), sehingga mencapai akurasi hingga 99,95% pada NSL-KDD dan mendekati sempurna pada berbagai kelas serangan di dataset CICIDS2017 (ScienceDirect, 2025a). Studi keempat menyoroti tantangan zero-day secara khusus, dengan menggabungkan autoencoder sebagai detektor anomali bersama Random Forest (model Random Forest-AE), yang diuji pada dataset CIC-MalMem-2022 dan data yang sepenuhnya belum pernah dilihat model sebelumnya (unseen data), mencapai akurasi 99,9892% (PMC, 2025).

Gambar 4.2 membandingkan hasil akurasi dari keempat pendekatan tersebut.



Gambar 4.2 Perbandingan Akurasi Model pada Studi Kasus Deteksi Intrusi Dataset NSL-KDD (disintesis oleh penulis)

Pembelajaran dari Studi Kasus

Perbandingan pada Gambar 4.2 menunjukkan tiga pembelajaran penting. Pertama, model ensemble yang menggabungkan beberapa algoritma

secara konsisten mengungguli model tunggal, meskipun selisihnya relatif kecil dalam angka absolut. Kedua, penyetelan hyperparameter yang cermat (seperti pada model XGBoost teroptimasi) dapat meningkatkan performa secara signifikan dibandingkan konfigurasi baku. Ketiga, dan yang paling penting untuk konteks buku ini, model yang dirancang khusus untuk menghadapi data zero-day (Random Forest-AE) mampu mempertahankan akurasi yang sangat tinggi bahkan pada data yang sepenuhnya baru, sebuah kapabilitas yang esensial mengingat karakteristik ancaman siber modern yang terus berevolusi, sebagaimana dibahas pada Bab 1 dan Bab 2.

4.7 Implementasi Studi Kasus 1: Pseudocode Pipeline Ensemble Stacking

Untuk memperjelas bagaimana konsep pada Subbab 4.2 dan 4.3 diterapkan secara konkret, bagian ini menyajikan pseudocode implementasi pipeline deteksi intrusi berbasis ensemble stacking, merepresentasikan pendekatan yang digunakan pada studi ketiga di Subbab 4.6 (kombinasi Random Forest, XGBoost, dan Extra-Trees dengan meta-model regresi logistik).

Pseudocode 4.1 — Prapemrosesan dan Seleksi Fitur

```
INPUT: dataset_mentah (NSL-KDD, 41 fitur, label multi-kelas)
OUTPUT: X_train, X_test, y_train, y_test (data siap latih)

FUNGSI preprocessing(dataset_mentah):
    // 1. Encoding fitur kategorikal (protocol_type,
service, flag)
    UNTUK setiap fitur_kategorikal DALAM dataset_mentah:
        dataset_mentah[fitur_kategorikal] =
OneHotEncode(fitur_kategorikal)

    // 2. Normalisasi fitur numerik ke rentang [0,1]
```

```
    UNTUK setiap fitur_numerik DALAM dataset_mentah:
        dataset_mentah[fitur_numerik] =
MinMaxScale(fitur_numerik)

    // 3. Seleksi fitur dengan Recursive Feature Elimination
(RFE)
    model_dasar = RandomForestClassifier()
    fitur_terpilih = RFE(estimator=model_dasar,
n_fitur_target=25)
                         .fit(dataset_mentah.X,
dataset_mentah.y)
                         .fiturTerpilih()

    // 4. Split data latih dan uji (80:20, stratified)
    X_train, X_test, y_train, y_test = train_test_split(
        dataset_mentah[fitur_terpilih], dataset_mentah.y,
        test_size=0.2, stratify=dataset_mentah.y,
random_state=42)

    KEMBALIKAN X_train, X_test, y_train, y_test
```

Pseudocode 4.2 — Pelatihan Model Ensemble Stacking

```
INPUT: X_train, y_train
OUTPUT: model_stacking (model ensemble terlatih)

FUNGSI latih_ensemble_stacking(X_train, y_train):
    // 1. Definisikan model dasar (base learners)
    model_rf  = RandomForestClassifier(n_estimators=200,
max_depth=15)
    model_xgb = XGBoostClassifier(n_estimators=300,
learning_rate=0.05)
    model_et  = ExtraTreesClassifier(n_estimators=200)

    // 2. Latih setiap model dasar menggunakan 5-fold cross-
validation
```

```
    //    untuk menghasilkan out-of-fold predictions sebagai
fitur meta
    prediksi_meta = MatriksKosong(baris=len(X_train),
kolom=3)
    UNTUK setiap (idx_latih, idx_val) DALAM KFold(X_train,
k=5):
        UNTUK setiap (i, model) DALAM enumerate([model_rf,
model_xgb, model_et]):
            model.fit(X_train[idx_latih],
y_train[idx_latih])
            prediksi_meta[idx_val][i] =
model.predict_proba(X_train[idx_val])

    // 3. Latih ulang setiap model dasar pada seluruh data
latih
    model_rf.fit(X_train, y_train)
    model_xgb.fit(X_train, y_train)
    model_et.fit(X_train, y_train)

    // 4. Latih meta-model (regresi logistik) pada
prediksi_meta
    meta_model = LogisticRegression()
    meta_model.fit(prediksi_meta, y_train)

    model_stacking = { base: [model_rf, model_xgb,
model_et], meta: meta_model }
    KEMBALIKAN model_stacking
```

Pseudocode 4.3 — Prediksi dan Evaluasi

```
INPUT: model_stacking, X_test, y_test
OUTPUT: metrik_evaluasi (akurasi, presisi, recall, F1-score)

FUNGSI prediksi_dan_evaluasi(model_stacking, X_test,
y_test):
```

```
    // 1. Hasilkan prediksi probabilitas dari setiap model
dasar
    pred_rf  = model_stacking.base[0].predict_proba(X_test)
    pred_xgb = model_stacking.base[1].predict_proba(X_test)
    pred_et  = model_stacking.base[2].predict_proba(X_test)
    fitur_meta_test = Gabungkan(pred_rf, pred_xgb, pred_et)

    // 2. Prediksi akhir menggunakan meta-model
    y_pred = model_stacking.meta.predict(fitur_meta_test)

    // 3. Hitung metrik evaluasi berdasarkan confusion
matrix (Subbab 4.2.7)
    metrik_evaluasi = {
        akurasi:  HitungAkurasi(y_test, y_pred),
        presisi:  HitungPresisi(y_test, y_pred),
        recall:   HitungRecall(y_test, y_pred),
        f1_score: HitungF1(y_test, y_pred)
    }
    KEMBALIKAN metrik_evaluasi
```

Ketiga potongan pseudocode di atas merepresentasikan alur kerja lengkap yang telah diilustrasikan secara konseptual pada Gambar 4.1: prapemrosesan dan seleksi fitur (Pseudocode 4.1) berkorespondensi dengan tahap 2–3, pelatihan ensemble (Pseudocode 4.2) berkorespondensi dengan tahap 4, dan evaluasi (Pseudocode 4.3) berkorespondensi dengan tahap 5. Pendekatan out-of-fold prediction pada Pseudocode 4.2 penting untuk dicermati: meta-model dilatih menggunakan prediksi dari data yang tidak dilihat langsung oleh model dasar selama fold tersebut, mencegah kebocoran informasi (data leakage) yang dapat menghasilkan estimasi performa yang terlalu optimistis.

4.8 Implementasi Studi Kasus: Deteksi Zero-Day dengan Random Forest-Autoencoder

Melengkapi pembahasan Studi Kasus 1, bagian ini menyajikan pseudocode untuk pendekatan Random Forest-AE yang secara khusus dirancang menghadapi ancaman zero-day, mengombinasikan autoencoder (Subbab 5.2, dibahas mendalam pada Bab 5) sebagai detektor anomali dengan Random Forest sebagai classifier utama.

Pseudocode 4.4 — Deteksi Zero-Day dengan Random Forest-Autoencoder

```
INPUT: X_train (data normal & serangan dikenal),
X_test_unseen (data zero-day)
OUTPUT: prediksi_akhir (label prediksi untuk setiap sampel
uji)

FUNGSI latih_rf_autoencoder(X_train, y_train):
    // 1. Latih autoencoder HANYA pada data trafik normal
    X_normal = FilterKelasNormal(X_train, y_train)
    autoencoder = BangunAutoencoder(
        dim_input=X_normal.shape[1], dim_latent=16)
    autoencoder.fit(X_normal, X_normal, epochs=50)   //
rekonstruksi diri sendiri

    // 2. Tentukan ambang batas (threshold) error
rekonstruksi
    error_rekonstruksi = HitungMSE(X_normal,
autoencoder.predict(X_normal))
    threshold = Percentile(error_rekonstruksi, 95)

    // 3. Latih Random Forest pada seluruh data berlabel
(normal + serangan dikenal)
    model_rf = RandomForestClassifier(n_estimators=300)
    model_rf.fit(X_train, y_train)
```

```
    KEMBALIKAN autoencoder, threshold, model_rf

FUNGSI deteksi(autoencoder, threshold, model_rf,
X_test_unseen):
    prediksi_akhir = []
    UNTUK setiap sampel x DALAM X_test_unseen:
        error = HitungMSE(x, autoencoder.predict(x))
        JIKA error > threshold:
            // Anomali signifikan -> kemungkinan pola belum
dikenal (zero-day)
            label = "ANOMALI_TERINDIKASI_ZERO_DAY"
        LAINNYA:
            // Pola dikenali -> klasifikasikan dengan Random
Forest
            label = model_rf.predict(x)
        prediksi_akhir.tambahkan(label)
    KEMBALIKAN prediksi_akhir
```

Logika inti pada Pseudocode 4.4 terletak pada mekanisme dua lapis: autoencoder berperan sebagai "penjaga gerbang" yang mendeteksi apakah suatu sampel menyimpang signifikan dari pola normal (diukur melalui mean squared error/MSE rekonstruksi), sementara Random Forest menangani klasifikasi rinci untuk sampel yang masih berada dalam batas pola yang telah dikenal. Pendekatan berlapis inilah yang menjelaskan mengapa model Random Forest-AE pada studi keempat (Subbab 4.6) mampu mempertahankan akurasi tinggi (99,9892%) bahkan pada data yang sepenuhnya belum pernah dilihat sebelumnya, ancaman baru tidak dipaksakan masuk ke salah satu kategori yang telah dipelajari, melainkan ditandai secara eksplisit sebagai anomali yang memerlukan investigasi lebih lanjut oleh analis.

4.9 Studi Kasus 2: Deteksi Anomali Transaksi Perbankan secara Real-Time

Konteks dan Permasalahan

Studi kasus kedua mengangkat penerapan machine learning tak terawasi (unsupervised) di luar konteks trafik jaringan, yaitu pada deteksi anomali transaksi perbankan. Sebuah penelitian tahun 2026 mengevaluasi kinerja platform Elastic Machine Learning, perangkat deteksi anomali tak terawasi tingkat industri dalam mengidentifikasi perilaku transaksi yang mencurigakan pada tiga skenario: keterkaitan tidak wajar antara satu akun dengan banyak alamat IP, lonjakan transaksi lintas negara dalam rentang waktu singkat, dan nilai transaksi yang jauh melampaui pola historis nasabah (MDPI, 2026).

Pendekatan dan Hasil

Berbeda dengan studi kasus pertama yang menggunakan pembelajaran terawasi dengan label eksplisit, pendekatan pada studi kasus ini bersifat tak terawasi, model mempelajari "pola normal" perilaku transaksi setiap akun, kemudian menandai penyimpangan signifikan sebagai potensi anomali tanpa memerlukan contoh data fraud yang telah diberi label sebelumnya. Hasil penelitian menunjukkan bahwa platform tersebut secara konsisten berhasil mendeteksi pola anomali pada ketiga skenario pengujian, dengan menandai perilaku mencurigakan secara real-time pada jendela waktu terjadinya kecurangan (MDPI, 2026).

Pembelajaran dari Studi Kasus

Studi kasus ini menggarisbawahi keunggulan pendekatan unsupervised untuk domain yang datanya terus berevolusi dan sulit diberi label secara lengkap, seperti pola kecurangan finansial yang terus berubah taktik.

Pendekatan ini melengkapi studi kasus pertama: jika deteksi intrusi jaringan pada NSL-KDD mengandalkan data berlabel historis yang kaya, deteksi anomali transaksi perbankan menunjukkan bagaimana machine learning tak terawasi dapat tetap efektif meski pola "normal" nasabah sangat bervariasi antarindividu dan berubah dari waktu ke waktu.

Pseudocode Implementasi: Skoring Anomali Berbasis K-Means

Untuk mengilustrasikan secara konkret prinsip kerja deteksi anomali tak terawasi yang mendasari studi kasus ini, berikut disajikan pseudocode sederhana berbasis K-Means (Subbab 4.2.5) yang diterapkan pada skenario deteksi transaksi mencurigakan.

Pseudocode 4.5 — Skoring Anomali Transaksi Perbankan Berbasis K-Means

```
INPUT: transaksi_historis (data transaksi normal per akun,
tanpa label fraud)
OUTPUT: skor_anomali (untuk setiap transaksi baru)

FUNGSI latih_profil_normal(transaksi_historis):
    // 1. Ekstraksi fitur perilaku per transaksi
    fitur = EkstrakFitur(transaksi_historis,
        kolom=[nilai_transaksi, jumlah_ip_unik_24jam,
               jumlah_negara_tujuan_1jam,
waktu_sejak_login])
    fitur = StandarisasiZScore(fitur)

    // 2. Bangun profil "normal" dengan K-Means (K
ditentukan via elbow method)
    model_kmeans = KMeans(k=5)
    model_kmeans.fit(fitur)

    KEMBALIKAN model_kmeans, parameter_standarisasi
```

```
FUNGSI skor_anomali(model_kmeans, parameter_standarisasi,
transaksi_baru):
    fitur_baru = EkstrakFitur(transaksi_baru)
    fitur_baru = Terapkan(parameter_standarisasi,
fitur_baru)

    // Jarak ke centroid terdekat sebagai indikator anomali
(Subbab 4.2.5)
    centroid_terdekat =
model_kmeans.centroidTerdekat(fitur_baru)
    skor = JarakEuclidean(fitur_baru, centroid_terdekat)

    JIKA skor > ambang_batas_dinamis(model_kmeans):
        KEMBALIKAN { status: "ANOMALI", skor: skor,
                     rekomendasi: "Tandai untuk verifikasi
tambahan" }
    LAINNYA:
        KEMBALIKAN { status: "NORMAL", skor: skor }
```

Perlu dicatat bahwa pseudocode di atas menyederhanakan pendekatan yang sesungguhnya diterapkan platform Elastic Machine Learning pada studi kasus ini, yang menggunakan model probabilistik yang lebih canggih untuk menangani multidimensi waktu dan populasi (population analysis) secara simultan (MDPI, 2026). Namun, prinsip inti tetap sama: mengukur seberapa jauh sebuah observasi baru menyimpang dari profil perilaku yang telah dipelajari, sejalan dengan formulasi K-Means pada Subbab 4.2.5.

Catatan: Praktik Terbaik Membangun IDS Berbasis Machine Learning

- *Gunakan kombinasi beberapa teknik seleksi fitur (RFE, Mutual Information, Lasso) untuk menghindari bias satu metode, sebagaimana ditunjukkan efektif dalam meningkatkan akurasi*

deteksi pada dataset CICIDS2017 dan NSL-KDD (ScienceDirect, 2025a).

- *Pendekatan ensemble/stacking (menggabungkan Random Forest, XGBoost, dan Extra-Trees) terbukti secara konsisten mengungguli model tunggal dalam berbagai studi terbaru (ScienceDirect, 2025a; arXiv, 2025).*
- *Untuk menghadapi ancaman zero-day, kombinasikan classifier supervised dengan autoencoder sebagai lapisan deteksi anomali tambahan, pendekatan ini terbukti mempertahankan akurasi tinggi bahkan pada data yang sepenuhnya belum pernah dilihat model sebelumnya (PMC, 2025).*
- *Selalu tangani ketidakseimbangan kelas (class imbalance) menggunakan teknik seperti SMOTE, mengingat data serangan pada dunia nyata jauh lebih jarang dibandingkan data trafik normal (ScienceDirect, 2025b).*
- *Perlakukan model machine learning untuk IDS sebagai sistem yang hidup: lakukan pelatihan ulang (retraining) secara berkala menggunakan data insiden terbaru, karena pola ancaman terus berevolusi sebagaimana dibahas pada Bab 1.*

Catatan: Keterbatasan yang Perlu Diwaspadai

- *Model machine learning untuk IDS rentan terhadap serangan adversarial, masukan yang secara sengaja dirancang untuk mengelabui model agar salah klasifikasi, sambil tetap tampak sebagai trafik yang sah (ScienceDirect, 2025b).*
- *Akurasi tinggi pada dataset benchmark (mis. NSL-KDD, CICIDS2017) tidak selalu berarti performa yang sama baiknya pada trafik dunia*

nyata, karena karakteristik jaringan produksi dapat sangat berbeda dari data pelatihan (ScienceDirect, 2025b).

- *Model yang kompleks seperti ensemble dan deep neural network cenderung sulit diinterpretasikan (black box), sehingga integrasi dengan explainable AI (dibahas pada Bab 7) menjadi penting agar analis keamanan dapat mempercayai dan memvalidasi keputusan model.*

4.10 Rangkuman

Deteksi berbasis machine learning terbagi menjadi pendekatan supervised (memerlukan data berlabel) dan unsupervised (mengenali pola tanpa label), yang masing-masing cocok untuk konteks permasalahan berbeda. Pembangunan IDS berbasis machine learning mengikuti alur kerja sistematis: pengumpulan data, prapemrosesan, seleksi fitur, pelatihan model, evaluasi, hingga penerapan dan pemantauan berkelanjutan. Algoritma ensemble seperti Random Forest dan XGBoost secara konsisten menunjukkan performa terbaik pada berbagai dataset benchmark seperti NSL-KDD, CICIDS2017, dan UNSW-NB15. Studi kasus deteksi intrusi jaringan menunjukkan bahwa kombinasi classifier supervised dengan autoencoder efektif menjaga akurasi tinggi bahkan pada data zero-day yang belum pernah terlihat model. Studi kasus deteksi anomali transaksi perbankan menggambarkan bagaimana pendekatan unsupervised tetap efektif untuk domain dengan pola "normal" yang sangat bervariasi dan terus berubah. Model machine learning untuk keamanan siber tetap memiliki keterbatasan, termasuk kerentanan terhadap serangan adversarial dan

tantangan interpretabilitas yang perlu diatasi dengan pendekatan explainable AI.

Daftar Pustaka

ScienceDirect. (2025a). Intrusion Detection System with Customized Machine Learning Techniques for NSL-KDD Dataset.

PMC (National Center for Biotechnology Information). (2025). An Intrusion Detection Model to Detect Zero-Day Attacks in Unseen Data Using Machine Learning.

arXiv. (2025). Binary and Multiclass Cyberattack Classification on GeNIS Dataset. arXiv:2511.08660.

ScienceDirect. (2025b). Enhancing IDS Performance through a Comparative Analysis of Random Forest, XGBoost, and Deep Neural Networks.

Nature Scientific Reports. (2025). Smart Deep Learning Model for Enhanced IoT Intrusion Detection.

MDPI. (2026). Detecting Cyber Fraud in Banking Transactions via Machine Learning Techniques: Implications for Financial Stability.

BAB 5

Deep Learning dalam Analisis Malware dan Traffic Jaringan

5.1 Pendahuluan

Jika Bab 4 berfokus pada machine learning klasik yang memerlukan rekayasa fitur manual, bab ini membahas bagaimana deep learning memungkinkan sistem keamanan siber mempelajari representasi fitur secara otomatis langsung dari data mentah. Kemampuan ini sangat relevan untuk dua tugas yang menjadi fokus bab ini: klasifikasi malware dan analisis pola traffic jaringan, dua domain yang datanya sering kali berdimensi tinggi, kompleks, dan sulit direkayasa fiturnya secara manual.

5.2 Landasan Fundamental Deep Learning

Sebelum membahas arsitektur spesifik seperti CNN dan LSTM, bagian ini menjabarkan fondasi matematis yang mendasari seluruh model deep learning: bagaimana neuron buatan bekerja, bagaimana jaringan belajar melalui backpropagation, serta formulasi matematis dari operasi konvolusi, rekursi, dan mekanisme attention yang menjadi inti pembahasan bab ini.

5.2.1 Neuron Buatan dan Fungsi Aktivasi

Unit dasar seluruh neural network adalah neuron buatan (perceptron), yang menghitung kombinasi linear dari input dikalikan bobot, ditambah bias, kemudian melewatkan hasilnya melalui fungsi aktivasi non-linear:

$$a = \varphi(\Sigma_i\, w_i x_i + b)$$

dengan x_i adalah nilai input, w_i adalah bobot yang dipelajari, b adalah bias, dan ϕ adalah fungsi aktivasi. Fungsi aktivasi yang paling umum digunakan pada arsitektur modern adalah ReLU (Rectified Linear Unit): ϕ(z) = max(0, z), karena kesederhanaannya secara komputasi dan efektivitasnya dalam mengatasi masalah vanishing gradient dibandingkan fungsi sigmoid klasik. Non-linearitas inilah yang memungkinkan neural network mempelajari pola yang jauh lebih kompleks dibandingkan model linear seperti regresi logistik.

5.2.2 Backpropagation dan Fungsi Kerugian

Neural network belajar melalui proses backpropagation: menghitung gradien fungsi kerugian terhadap setiap bobot dalam jaringan, kemudian memperbarui bobot tersebut ke arah yang meminimalkan kerugian, menggunakan aturan pembaruan gradient descent:

$$w_i \leftarrow w_i - \eta \cdot (\partial L / \partial w_i)$$

dengan η adalah learning rate dan L adalah fungsi kerugian (misalnya cross-entropy loss untuk klasifikasi). Proses ini diulang secara iteratif pada seluruh data pelatihan (dalam batch-batch kecil, dikenal sebagai mini-batch gradient descent) hingga model konvergen pada nilai kerugian yang minimal. Kedalaman (jumlah lapisan) pada arsitektur deep learning memungkinkan jaringan mempelajari representasi fitur secara hierarkis lapisan awal mempelajari pola sederhana, lapisan lebih dalam mengombinasikan pola tersebut menjadi representasi yang semakin abstrak dan kompleks.

5.2.3 Operasi Konvolusi pada CNN

Convolutional Neural Network bekerja dengan menggeser filter (kernel) berukuran kecil ke seluruh area input, menghitung hasil kali titik (dot product) pada setiap posisi untuk menghasilkan feature map. Untuk input dua dimensi I dan kernel K, operasi konvolusi dirumuskan sebagai:

$$S(i,j) = \Sigma_m \Sigma_n I(i+m, j+n) \cdot K(m,n)$$

Setiap filter dilatih untuk mendeteksi pola visual tertentu, pada konteks analisis malware yang dibahas pada bab ini, filter pada lapisan awal dapat mendeteksi tekstur biner sederhana, sementara filter pada lapisan lebih dalam mendeteksi pola struktural yang lebih kompleks dan spesifik terhadap keluarga malware tertentu. Operasi pooling (umumnya max pooling) kemudian diterapkan untuk mereduksi dimensi feature map sekaligus mempertahankan fitur yang paling menonjol: MaxPool mengambil nilai maksimum pada setiap area kecil dari feature map, mengurangi ukuran representasi sekaligus memberikan invariansi terhadap pergeseran posisi kecil pada input.

5.2.4 Recurrent Neural Network dan LSTM

RNN dirancang untuk memproses data sekuensial dengan mempertahankan hidden state yang membawa informasi dari langkah waktu sebelumnya:

$$h_t = \varphi(W_h \cdot h_{t-1} + W_x \cdot x_t + b)$$

dengan h_t adalah hidden state pada waktu t, dan x_t adalah input pada waktu t. Namun, RNN klasik mengalami masalah vanishing gradient pada sekuens panjang, sehingga informasi dari langkah waktu awal cenderung "terlupakan". Long Short-Term Memory (LSTM) mengatasi keterbatasan ini

melalui mekanisme gerbang (gate) yang secara eksplisit mengontrol aliran informasi:

$$f_t = \sigma(Wf \cdot [h_{t-1}, x_t] + bf) \quad \text{— gerbang lupa (forget gate)}$$
$$i_t = \sigma(Wi \cdot [h_{t-1}, x_t] + bi) \quad \text{— gerbang masuk (input gate)}$$
$$o_t = \sigma(Wo \cdot [h_{t-1}, x_t] + bo) \quad \text{— gerbang keluar (output gate)}$$

Gerbang lupa (f_t) menentukan informasi mana dari cell state sebelumnya yang perlu dibuang, gerbang masuk (i_t) menentukan informasi baru mana yang perlu disimpan, dan gerbang keluar (o_t) menentukan bagian mana dari cell state yang akan digunakan untuk menghasilkan hidden state saat ini. Melalui mekanisme ini, LSTM mampu mempertahankan dependensi jangka Panjang, kapabilitas yang esensial untuk menganalisis sekuens panjang seperti rangkaian pemanggilan API selama eksekusi program, sebagaimana dibahas pada studi kasus di Subbab 5.4.

5.2.5 Autoencoder dan Mekanisme Attention

Autoencoder terdiri dari dua komponen: encoder yang memampatkan input x menjadi representasi laten berdimensi rendah z = f(x), dan decoder yang berupaya merekonstruksi kembali input asli dari representasi laten tersebut, $\hat{x} = g(z)$. Model dilatih untuk meminimalkan kerugian rekonstruksi:

$$L\_rekonstruksi = \|x - \hat{x}\|^2$$

Karena dilatih hanya pada data normal, autoencoder akan menghasilkan error rekonstruksi yang rendah untuk data normal, namun error yang tinggi untuk data anomali yang polanya belum pernah dipelajari. Prinsip inilah yang mendasari deteksi anomali berbasis autoencoder yang telah disinggung pada Bab 4 dan akan diterapkan kembali pada Pseudocode 5.2 di bab ini.

Mekanisme attention, yang menjadi komponen kunci pada studi kasus Subbab 5.4, menghitung bobot relevansi setiap elemen input terhadap output yang akan dihasilkan, memungkinkan model secara adaptif "memfokuskan" perhatian pada bagian input yang paling informatif:

$$\alpha_t = softmax(score(h_t, h)) \qquad context = \Sigma_t \, \alpha_t \cdot h_t$$

dengan α_t adalah bobot attention untuk langkah waktu t, dihitung menggunakan fungsi skor (umumnya dot product atau small feed-forward network) yang mengukur relevansi hidden state h_t terhadap konteks keseluruhan. Vektor context yang dihasilkan. Kombinasi terbobot dari seluruh hidden state memberikan representasi yang lebih kaya dibandingkan hanya mengandalkan hidden state pada langkah waktu terakhir, sekaligus menawarkan interpretabilitas tambahan karena bobot α_t menunjukkan bagian input mana yang paling memengaruhi keputusan model.

5.2.6 Contoh Numerik: Operasi Konvolusi pada Potongan Data Biner

Untuk memperjelas formulasi pada Subbab 5.2.3, berikut contoh numerik sederhana. Misalkan sebuah potongan kecil representasi citra dari berkas biner (setelah konversi menggunakan Pseudocode 5.1) berukuran 4×4 piksel, dan kernel deteksi tepi (edge detector) berukuran 2×2 berikut diterapkan:

Tabel 5.2 Ilustrasi Numerik Operasi Konvolusi

Komponen	Nilai (Contoh)
Potongan citra input (4×4)	[[12,45,78,23],[90,34,56,120],[200,15,88,44],[10,60,130,25]]
Kernel/filter (2×2)	[[1,-1],[1,-1]]

Hasil konvolusi posisi (0,0)	(12×1)+(45×−1)+(90×1)+(34×−1) = 3
Hasil konvolusi posisi (0,1)	(45×1)+(78×−1)+(34×1)+(56×−1) = −55
Interpretasi	Nilai absolut besar (negatif/positif) mengindikasikan tepi/perubahan tajam pola byte — sering berkorelasi dengan header atau instruksi kode tereksekusi pada malware

Proses ini diulang dengan menggeser kernel ke seluruh area citra (operasi sliding window), menghasilkan feature map yang menonjolkan pola perubahan nilai byte yang tajam. Pada CNN sesungguhnya, ratusan hingga ribuan kernel semacam ini dipelajari secara otomatis melalui backpropagation (Subbab 5.2.2), masing-masing belajar mendeteksi pola berbeda yang relevan untuk membedakan keluarga malware satu dengan lainnya, tanpa perlu didefinisikan secara manual oleh insinyur.

5.2.7 Kompleksitas Komputasi dan Trade-off Arsitektur

Pemilihan arsitektur deep learning tidak lepas dari pertimbangan kompleksitas komputasi, yang secara langsung memengaruhi kelayakan penerapan real-time sebagaimana ditekankan pada studi kasus Subbab 5.5. Tabel 5.3 merangkum karakteristik kompleksitas relatif dari komponen arsitektur yang telah dibahas.

Tabel 5.3 Perbandingan Kompleksitas Komputasi Komponen Arsitektur

Komponen	Kompleksitas Relatif	Implikasi Praktis
Lapisan Konvolusi (CNN)	Sedang. Dapat diparalelkan secara efisien pada GPU	Cocok untuk pemrosesan batch besar; latensi rendah pada perangkat keras modern

Lapisan LSTM	Tinggi. Komputasi sekuensial, sulit diparalelkan penuh	Menjadi bottleneck utama pada sekuens panjang; motivasi penggunaan arsitektur transformer pada riset lebih lanjut
Lapisan Attention	Sedang-Tinggi. Bergantung pada panjang sekuens (kompleksitas kuadratik)	Efektif untuk sekuens pendek-menengah (log/API call); perlu optimasi untuk sekuens sangat panjang
Autoencoder	Rendah-Sedang. Bergantung pada ukuran lapisan laten	Relatif ringan, cocok untuk deteksi anomali dengan sumber daya terbatas

Trade-off ini menjelaskan mengapa studi kasus pada Subbab 5.5 secara khusus menekankan pencapaian latensi di bawah 35 milidetik sebagai kontribusi penting bukan sekadar akurasi tinggi yang mudah dicapai dengan menambah kompleksitas model, melainkan akurasi tinggi yang dicapai tanpa mengorbankan kelayakan operasional real-time, sebuah keseimbangan yang tidak selalu mudah dicapai mengingat karakteristik komputasi LSTM dan attention yang cenderung lebih berat dibandingkan CNN murni.

5.3 Arsitektur Deep Learning untuk Keamanan Siber

Beberapa arsitektur deep learning telah terbukti efektif dalam konteks keamanan siber, masing-masing dengan kekuatan yang berbeda tergantung karakteristik data yang dihadapi. Tabel 5.1 merangkum empat arsitektur utama yang paling banyak diadopsi pada penelitian dan implementasi terbaru.

Tabel 5.1 Arsitektur Deep Learning untuk Analisis Malware dan Traffic Jaringan

Arsitektur	Jenis Data yang Ditangani	Kekuatan Utama	Contoh Kinerja Terbaru
CNN (Convolutional Neural Network)	Citra biner malware, representasi spasial trafik	Ekstraksi fitur spasial otomatis; efektif pada pola visual berulang	Deteksi malware real-time pada infrastruktur cloud (American Journal of Engineering & Technology, 2025)
RNN / LSTM	Data sekuensial: log sistem, urutan panggilan API, aliran trafik	Menangkap dependensi temporal jangka panjang	Klasifikasi keluarga malware berbasis opcode dan API call (Springer Nature, 2024)
Autoencoder	Data tanpa label untuk deteksi anomali	Belajar merekonstruksi pola normal; anomali dikenali dari error rekonstruksi	Deteksi intrusi jaringan tanpa memerlukan data serangan berlabel (Binbusayyis & Vaiyapuri, 2021)
Hybrid CNN-LSTM + Attention	Kombinasi data spasial dan sekuensial dengan bobot perhatian	Menggabungkan keunggulan CNN dan LSTM; attention layer meningkatkan fokus pada fitur penting	94,8–97,5% akurasi pada NSL-KDD dan Bot-IoT, latensi inferensi di bawah 35 milidetik (PMC / Nature Scientific Reports, 2025)

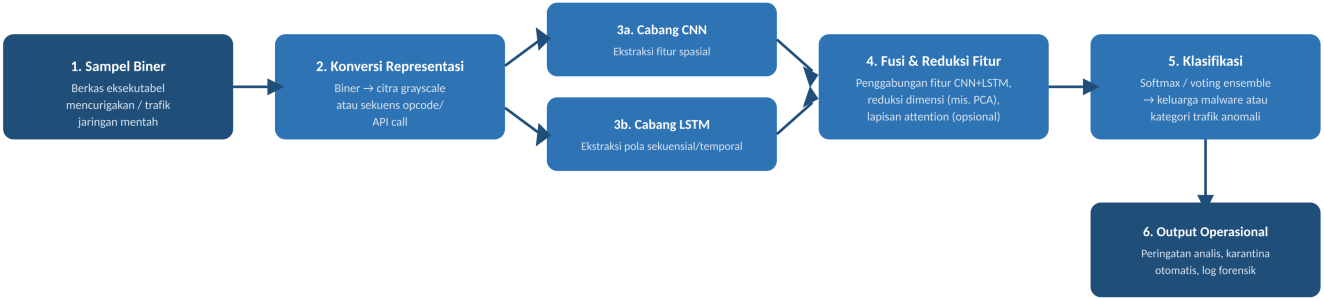
5.4 Prinsip Kerja: Dari Biner Mentah menuju Klasifikasi

Pendekatan yang banyak digunakan dalam analisis malware berbasis deep learning adalah mengonversi berkas biner menjadi representasi citra

grayscale, di mana setiap byte kode diperlakukan sebagai nilai piksel. Representasi ini memungkinkan CNN mengenali pola visual berulang yang khas pada varian malware tertentu, layaknya mengenali pola pada citra digital biasa. Untuk menangkap aspek temporal, misalnya urutan pemanggilan fungsi sistem (API call) atau opcode selama eksekusi program arsitektur LSTM dikombinasikan untuk mempelajari dependensi jangka panjang dalam sekuens tersebut.

Gambar 5.1 mengilustrasikan alur kerja pendekatan hybrid CNN-LSTM secara menyeluruh, mulai dari sampel biner mentah hingga keluaran operasional berupa peringatan atau tindakan karantina otomatis.

Gambar 5.1 Alur Analisis Malware Berbasis Deep Learning (Pendekatan CNN-LSTM)



Sumber: disintesis oleh penulis dari Wireless Personal Communications, Springer (2024); PMC/Nature Scientific Reports (2025); American Journal of Engineering & Technology (2025).

Gambar 5.1 Alur Analisis Malware Berbasis Deep Learning (Pendekatan CNN-LSTM) (disintesis oleh penulis)

5.5 Studi Kasus: Deteksi Intrusi Jaringan Real-Time dengan Model Attention-CNN-LSTM

Konteks dan Permasalahan

Sebuah penelitian yang dipublikasikan pada Scientific Reports (Nature) tahun 2025 mengangkat permasalahan klasik dalam deteksi intrusi berbasis deep learning: bagaimana meningkatkan akurasi deteksi tanpa mengorbankan kecepatan inferensi yang dibutuhkan untuk operasi keamanan real-time. Banyak model deep learning yang akurat namun terlalu lambat untuk diterapkan pada trafik jaringan berkecepatan tinggi, sementara model yang cepat cenderung kurang akurat dalam mengenali pola serangan yang kompleks (PMC / Nature Scientific Reports, 2025).

Pendekatan

Peneliti mengembangkan model Attention-CNN-LSTM yang menggabungkan tiga komponen: lapisan CNN untuk ekstraksi fitur spasial dari data trafik, lapisan LSTM untuk menangkap pola sekuensial dari waktu ke waktu, dan mekanisme attention yang secara adaptif memberi bobot lebih besar pada fitur-fitur yang paling relevan terhadap indikasi serangan. Model ini diuji pada dua dataset benchmark yang telah dibahas pada Bab 4: NSL-KDD dan Bot-IoT.

Hasil dan Pembelajaran

Model yang diusulkan mencapai akurasi 94,8–97,5% pada kedua dataset, dengan peningkatan signifikan pada Matthews Correlation Coefficient (MCC) dan F1-score dibandingkan model dasar tanpa attention. Yang secara khusus relevan untuk penerapan praktis, penelitian ini menunjukkan bahwa arsitektur tersebut mendukung inferensi real-time dengan latensi di bawah 35 milidetik. Cukup cepat untuk diterapkan pada trafik jaringan produksi.

Studi ablasi (ablation study) yang dilakukan peneliti mengonfirmasi bahwa lapisan attention secara khusus memberikan kontribusi besar terhadap peningkatan performa keseluruhan, membuktikan bahwa penambahan mekanisme fokus adaptif bukan sekadar kompleksitas tambahan, melainkan peningkatan yang secara nyata terukur (PMC / Nature Scientific Reports, 2025).

Studi kasus ini memberikan pembelajaran penting bagi buku ini: peningkatan akurasi deep learning tidak harus mengorbankan kecepatan operasional, asalkan arsitektur dirancang secara cermat. Sebuah prinsip yang akan kembali relevan ketika membahas integrasi Graph Neural Networks pada Bab 6 dan sistem IDS/IPS real-time pada Bab 8.

Catatan: Praktik Terbaik Deep Learning untuk Keamanan Siber

- *Konversi data biner menjadi representasi citra atau sekuens memungkinkan pemanfaatan arsitektur CNN/LSTM yang telah matang tanpa memerlukan rekayasa fitur manual yang ekstensif (American Journal of Engineering & Technology, 2025; Springer Nature, 2024).*
- *Mekanisme attention terbukti secara konsisten meningkatkan baik akurasi maupun interpretabilitas model, karena bobot perhatian dapat menunjukkan fitur mana yang paling memengaruhi keputusan klasifikasi (PMC / Nature Scientific Reports, 2025).*
- *Untuk skenario tanpa data berlabel yang memadai, autoencoder tetap menjadi pilihan kuat untuk deteksi anomali berbasis rekonstruksi, sejalan dengan prinsip yang telah dibahas pada Bab 4 (Binbusayyis & Vaiyapuri, 2021).*

- *Pertimbangkan trade-off antara akurasi dan latensi sejak tahap perancangan arsitektur, khususnya untuk kasus penggunaan yang menuntut deteksi real-time pada trafik jaringan berkecepatan tinggi.*

5.6 Implementasi: Pseudocode Pipeline Deep Learning untuk Analisis Malware

Bagian ini menyajikan pseudocode implementasi menyeluruh yang menggabungkan konsep-konsep pada Subbab 5.2, merepresentasikan bagaimana pipeline hybrid CNN-LSTM dengan attention (Subbab 5.4–5.5) dibangun secara praktis, mulai dari representasi data hingga inferensi model terlatih.

Tahap 1: Representasi Data — Konversi Biner menjadi Citra

Pseudocode 5.1 — Konversi Berkas Biner menjadi Citra Grayscale

```
INPUT: berkas_biner (file .exe/.dll mencurigakan)
OUTPUT: citra_malware (matriks 2D nilai piksel 0-255)

FUNGSI konversi_biner_ke_citra(berkas_biner):
    byte_stream = BacaByteMentah(berkas_biner)     // setiap
byte = 0-255

    // Tentukan lebar citra secara adaptif berdasarkan
ukuran berkas
    lebar = TentukanLebarAdaptif(len(byte_stream))
    tinggi = CeilDiv(len(byte_stream), lebar)

    // Bentuk ulang (reshape) byte stream menjadi matriks 2D
    citra_malware = ReshapeDenganPadding(
```

```
        byte_stream, bentuk=(tinggi, lebar),
nilai_padding=0)

    KEMBALIKAN citra_malware   // setiap piksel
merepresentasikan 1 byte kode
```

Tahap 2: Arsitektur dan Pelatihan Model Hybrid

Pseudocode 5.2 — Arsitektur dan Pelatihan Model Attention-CNN-LSTM

```
INPUT: X_train (citra malware + sekuens API call), y_train
(label keluarga malware)
OUTPUT: model_hybrid (model terlatih)

FUNGSI bangun_model_hybrid(dim_citra, panjang_sekuens):
    // Cabang CNN – ekstraksi fitur spasial dari citra biner
    cabang_cnn = Sequential([
        Conv2D(filters=32, kernel_size=3, aktivasi='relu'),
        MaxPooling2D(pool_size=2),
        Conv2D(filters=64, kernel_size=3, aktivasi='relu'),
        MaxPooling2D(pool_size=2),
        Flatten()
    ])

    // Cabang LSTM – ekstraksi fitur sekuensial dari urutan
API call
    cabang_lstm = Sequential([
        Embedding(vocab_size=API_VOCAB_SIZE, dim_output=64),
        LSTM(unit=128, return_sequences=BENAR)
    ])

    // Lapisan Attention (Subbab 5.2.5) diterapkan pada
output LSTM
```

```
    attention_output, bobot_attention =
AttentionLayer(cabang_lstm.output)

    // Fusi fitur CNN + LSTM-Attention, lalu klasifikasi
akhir
    fitur_gabungan = Concatenate([cabang_cnn.output,
attention_output])
    fitur_gabungan = Dense(unit=64,
aktivasi='relu')(fitur_gabungan)
    output = Dense(unit=jumlah_kelas,
aktivasi='softmax')(fitur_gabungan)

    model_hybrid = Model(input=[cabang_cnn.input,
cabang_lstm.input],
                         output=output)
    model_hybrid.compile(optimizer='adam',
                         loss='categorical_crossentropy')
    KEMBALIKAN model_hybrid

FUNGSI latih(model_hybrid, X_train, y_train):
    // Pelatihan menggunakan backpropagation (Subbab 5.2.2)
    model_hybrid.fit(X_train, y_train, epochs=30,
batch_size=64,
                     validation_split=0.2,
early_stopping=BENAR)
    KEMBALIKAN model_hybrid
```

Tahap 3: Inferensi Real-Time dan Interpretasi Attention

Pseudocode 5.3 — Inferensi Real-Time dengan Interpretasi Bobot Attention

```
INPUT: model_hybrid, sampel_baru (citra + sekuens API call)
OUTPUT: prediksi (label + tingkat keyakinan + penjelasan
attention)
```

```
FUNGSI inferensi_real_time(model_hybrid, sampel_baru):
    waktu_mulai = WaktuSekarang()

    probabilitas, bobot_attention =
model_hybrid.predict_dengan_attention(
        sampel_baru)
    label_prediksi = ArgMax(probabilitas)
    keyakinan = Max(probabilitas)

    latensi = WaktuSekarang() - waktu_mulai   // target: <
35 ms (PMC / Nature Scientific Reports, 2025)

    // Identifikasi langkah API call dengan bobot attention
tertinggi
    // sebagai dasar penjelasan (menuju prinsip XAI pada Bab
7)
    api_paling_berpengaruh = TopK(bobot_attention, k=5)

    KEMBALIKAN {
        label: label_prediksi,
        keyakinan: keyakinan,
        latensi_ms: latensi,
        penjelasan: api_paling_berpengaruh
    }
```

Ketiga tahap pseudocode di atas mendemonstrasikan bagaimana landasan matematis pada Subbab 5.2 terwujud dalam implementasi praktis: representasi citra (Pseudocode 5.1) mengaplikasikan prinsip konversi data pada Subbab 5.3, arsitektur hybrid (Pseudocode 5.2) mengintegrasikan operasi konvolusi (Subbab 5.2.3), rekursi LSTM (Subbab 5.2.4), dan attention (Subbab 5.2.5) secara bersamaan, sementara inferensi real-time (Pseudocode 5.3) menunjukkan bagaimana bobot attention dapat diekstraksi bukan hanya untuk meningkatkan akurasi, tetapi juga sebagai

bentuk interpretabilitas awal yang menjembatani pembahasan menuju explainable AI pada Bab 7.

5.7 Rangkuman

Deep learning memungkinkan ekstraksi fitur otomatis dari data mentah, menjadikannya pilihan kuat untuk analisis malware dan traffic jaringan yang berdimensi tinggi dan kompleks. CNN unggul dalam menangkap pola spasial (representasi citra biner), LSTM unggul dalam menangkap pola sekuensial (log, opcode, API call), dan keduanya sering dikombinasikan dalam arsitektur hybrid. Mekanisme attention terbukti meningkatkan akurasi sekaligus memberikan indikasi interpretabilitas tambahan pada model deep learning. Studi kasus model Attention-CNN-LSTM menunjukkan bahwa akurasi tinggi (94,8–97,5%) dapat dicapai bersamaan dengan latensi inferensi yang sangat rendah (<35 milidetik), membuktikan kelayakan deep learning untuk deteksi real-time.

Daftar Pustaka

American Journal of Engineering & Technology. (2025). Real-Time Malware Detection in Cloud Infrastructures Using Convolutional Neural Networks: A Deep Learning Framework for Enhanced Cybersecurity.

Springer Nature. (2024). Hybrid Deep Learning Approach Based on LSTM and CNN for Malware Detection. Wireless Personal Communications.

Binbusayyis, A., & Vaiyapuri, T. (2021). Unsupervised Deep Learning Approach for Network Intrusion Detection Combining Convolutional Autoencoder and One-Class SVM. Applied Intelligence.

PMC / Nature Scientific Reports. (2025). Deep Learning for Network Security: An Attention-CNN-LSTM Model for Accurate Intrusion Detection. Sci Rep 15, 21856.

BAB 6

Graph Neural Networks. Memodelkan Hubungan Antarentitas dalam Jaringan

6.1 Pendahuluan

Bab ini membahas secara mendalam pendekatan graph-based learning yang telah diperkenalkan pada taksonomi Bab 3, dengan fokus khusus pada Graph Neural Networks (GNN). Pendekatan ini menjadi landasan konseptual utama bagi studi kasus implementasi yang akan disajikan pada Bab 11, sehingga pemahaman yang kokoh terhadap prinsip kerja GNN pada bab ini akan sangat membantu dalam memahami rancangan sistem pada bagian akhir buku.

6.2 Mengapa Merepresentasikan Data Keamanan Siber sebagai Graf?

Pendekatan machine learning dan deep learning konvensional yang dibahas pada Bab 4 dan 5 umumnya memperlakukan setiap sampel data (baris trafik, satu berkas malware) sebagai entitas yang independen satu sama lain. Padahal, dalam realitas operasional jaringan, ancaman siber jarang muncul sebagai kejadian tunggal yang berdiri sendiri, sebuah serangan modern umumnya melibatkan rangkaian interaksi antarentitas: penyerang

mengeksploitasi satu host, kemudian bergerak secara lateral ke host lain, mengakses kredensial pengguna, dan akhirnya mencapai target bernilai tinggi seperti server basis data.

Graph Neural Networks dirancang khusus untuk menangkap pola relasional semacam ini. Dengan merepresentasikan host, pengguna, server, dan paket data sebagai node, serta interaksi di antaranya sebagai edge, GNN dapat mempelajari pola dari struktur relasi itu sendiri, sesuatu yang secara fundamental tidak dapat dilakukan oleh model berbasis tabel data konvensional.

6.3 Landasan Fundamental Teori Graf dan Graph Neural Network

Memahami GNN secara mendalam menuntut pemahaman terhadap dua fondasi: teori graf klasik yang mendasari representasi data, dan formulasi matematis spesifik yang membedakan GCN, GAT, dan GraphSAGE sebagaimana diperkenalkan pada Bab 3.

6.3.1 Representasi Matematis Graf

Secara formal, sebuah graf G didefinisikan sebagai pasangan G = (V, E), dengan V adalah himpunan node (entitas seperti host, pengguna, server) dan E adalah himpunan edge (relasi/interaksi antarentitas). Struktur graf dapat direpresentasikan secara komputasional melalui matriks ketetanggaan (adjacency matrix) A berukuran n×n, dengan n adalah jumlah node:

$$A[i,j] = 1 \text{ jika terdapat edge antara node } i \text{ dan node } j; 0 \text{ jika sebaliknya}$$

Setiap node juga dapat memiliki vektor fitur sendiri (misalnya jumlah paket terkirim, jenis protokol yang digunakan oleh suatu host), dikumpulkan dalam matriks fitur node X berukuran n×d, dengan d adalah dimensi fitur. Matriks derajat (degree matrix) D adalah matriks diagonal dengan D[i,i] menyatakan jumlah edge yang terhubung ke node i. Kombinasi A, D, dan X inilah yang menjadi masukan dasar bagi seluruh varian GNN yang dibahas pada bab ini.

6.3.2 Formulasi Graph Convolutional Network (GCN)

GCN, sebagaimana diperkenalkan oleh Kipf & Welling, mendefinisikan operasi konvolusi pada graf melalui aturan propagasi berlapis (layer-wise propagation rule):

$$H^{(l+1)} = \varphi(\tilde{D}^{(-1/2)} \cdot \tilde{A} \cdot \tilde{D}^{(-1/2)} \cdot H^{(l)} \cdot W^{(l)})$$

dengan $\tilde{A} = A + I$ adalah matriks ketetanggaan yang telah ditambahkan self-loop (agar setiap node turut mempertimbangkan fitur dirinya sendiri), $\tilde{D}$ adalah matriks derajat dari $\tilde{A}$, $H^{(l)}$ adalah representasi node pada lapisan ke-l (dengan $H^{(0)} = X$, matriks fitur awal), $W^{(l)}$ adalah matriks bobot yang dipelajari pada lapisan tersebut, dan ϕ adalah fungsi aktivasi non-linear (umumnya ReLU, sebagaimana dibahas pada Bab 5). Normalisasi $\tilde{D}^{(-1/2)} \cdot \tilde{A} \cdot \tilde{D}^{(-1/2)}$ memastikan skala fitur tetap stabil terlepas dari jumlah tetangga yang dimiliki setiap node. Tanpa normalisasi ini, node dengan derajat tinggi (mis. server pusat yang terhubung ke banyak host) akan mendominasi secara tidak proporsional.

6.3.3 Formulasi Graph Attention Network (GAT)

Berbeda dengan GCN yang memberi bobot seragam (dinormalisasi berdasarkan derajat) pada seluruh tetangga, GAT mempelajari bobot

attention berbeda untuk setiap pasangan node bertetangga. Koefisien attention antara node i dan tetangganya j dihitung sebagai:

$$e_{ij} = LeakyReLU(a^T \cdot [W \cdot h_i \,\|\, W \cdot h_j])$$

$$\alpha_{ij} = softmax_j(e_{ij}) = exp(e_{ij}) / \Sigma_{k \in N(i)} exp(e_{ik})$$

dengan W adalah matriks transformasi linear bersama, a adalah vektor bobot attention yang dipelajari, ‖ menyatakan operasi penggabungan (concatenation), dan N(i) adalah himpunan tetangga node i. Representasi node i pada lapisan berikutnya kemudian dihitung sebagai kombinasi terbobot dari representasi tetangganya:

$$h_i^{(l+1)} = \varphi(\Sigma_{j \in N(i)} \alpha_{ij} \cdot W \cdot h_j^{(l)})$$

Formulasi ini secara matematis identik dengan prinsip attention yang telah dibahas pada Bab 5 (Subbab 5.2.5), namun diterapkan pada konteks relasi antarnode dalam graf, bukan pada langkah waktu dalam sekuens. Kemampuan GAT memberi bobot berbeda pada setiap tetangga menjadikannya sangat relevan untuk keamanan siber: tetangga yang menunjukkan pola mencurigakan (misalnya host dengan riwayat pelanggaran keamanan) dapat diberi bobot lebih tinggi dibandingkan tetangga yang berperilaku normal.

6.3.4 Formulasi GraphSAGE

GraphSAGE (Sample and Aggregate) dirancang untuk skenario graf berskala sangat besar dan dinamis, di mana menghitung seluruh tetangga suatu node (sebagaimana pendekatan GCN klasik) menjadi tidak praktis secara komputasi. GraphSAGE mengambil sampel acak dari sejumlah tetatangga (bukan seluruhnya), kemudian mempelajari fungsi agregasi yang dapat digeneralisasi:

$$h\_N(i) = AGGREGATE(\{h_j, \forall j \in Sampel(N(i))\})$$

$$h_i^{(l+1)} = \varphi(W \cdot [h_i^{(l)} \| h\_N(i)])$$

Fungsi AGGREGATE dapat berupa operasi rata-rata (mean), maksimum (max-pooling), atau bahkan LSTM (memanfaatkan kembali prinsip pada Bab 5) yang memproses tetangga sebagai sekuens. Karena fungsi agregasi dipelajari secara independen dari identitas node tertentu, GraphSAGE mampu melakukan inferensi pada node yang belum pernah terlihat selama pelatihan (inductive learning). Kapabilitas krusial bagi jaringan yang terus bertumbuh dengan host dan pengguna baru, sebagaimana disinggung pada Tabel 6.1.

6.3.5 Prinsip Umum Message Passing

Ketiga formulasi di atas (GCN, GAT, GraphSAGE) sesungguhnya merupakan kasus khusus dari kerangka kerja message passing yang lebih umum, yang dapat dirumuskan dalam dua tahap pada setiap lapisan:

$$m_{ij}^{(l)} = MESSAGE(h_i^{(l)}, h_j^{(l)}, e_{ij}) \quad \text{— tahap 1: hitung pesan dari } j \text{ ke } i$$

$$h_i^{(l+1)} = UPDATE(h_i^{(l)}, AGGREGATE(\{m_{ij}^{(l)} : j \in N(i)\})) \quad \text{— tahap 2: perbarui representasi } i$$

Kerangka umum inilah yang diilustrasikan secara visual pada Gambar 6.1(b): setiap node mengumpulkan "pesan" dari tetangganya (tahap MESSAGE), mengagregasi pesan tersebut (tahap AGGREGATE, dapat berupa rata-rata sederhana pada GCN atau terbobot attention pada GAT), kemudian memperbarui representasinya sendiri (tahap UPDATE). Pemahaman terhadap kerangka umum ini penting karena hampir seluruh varian GNN yang dikembangkan dalam literatur, termasuk yang akan diterapkan pada studi kasus Bab 11. Pada dasarnya adalah variasi dari ketiga fungsi MESSAGE, AGGREGATE, dan UPDATE ini.

6.3.6 Contoh Numerik: Propagasi Satu Lapis GCN

Untuk memperjelas formulasi pada Subbab 6.3.2, berikut contoh numerik sederhana pada graf mini beranggotakan 3 node. Merepresentasikan tiga host dalam jaringan kecil, dengan Host A dan Host B saling terhubung, serta Host B dan Host C saling terhubung (Host A tidak terhubung langsung dengan Host C).

Tabel 6.2 Ilustrasi Numerik Propagasi Graph Convolutional Network

Komponen	Nilai (Contoh)
Matriks ketetanggaan A (3×3)	[[0,1,0],[1,0,1],[0,1,0]] — A menghubungkan Host A-B dan B-C
Matriks fitur awal X (3×2)	Host A=[1,0], Host B=[0,1], Host C=[1,1] (fitur: aktivitas login, transfer data)
Matriks Ã (dengan self-loop)	A + I = [[1,1,0],[1,1,1],[0,1,1]]
Representasi Host B setelah 1 lapis GCN	Agregasi terbobot dari fitur Host A, Host B, dan Host C sendiri (karena self-loop), menghasilkan representasi baru yang mencerminkan konteks kedua tetangganya
Interpretasi	Setelah satu lapis propagasi, representasi Host B tidak lagi hanya mencerminkan fiturnya sendiri, melainkan juga "mewarisi" informasi dari Host A dan Host C — inilah mekanisme yang memungkinkan GNN mendeteksi pola pergerakan lateral seperti diilustrasikan pada Gambar 6.1(a)

Setelah beberapa lapis propagasi (umumnya 2–3 lapis pada praktiknya), representasi setiap node mencakup informasi dari tetangga hingga beberapa "lompatan" (hops) jauhnya. Memungkinkan model mendeteksi pola yang melibatkan rangkaian relasi tidak langsung, seperti penyerang

yang bergerak dari Host A ke Host C melalui perantara Host B, meskipun Host A dan Host C sendiri tidak pernah berinteraksi langsung. Kemampuan menangkap relasi tidak langsung inilah yang menjadi pembeda fundamental GNN dari model berbasis tabel data konvensional yang dibahas pada Bab 4, yang hanya dapat mengevaluasi setiap baris data secara independen.

6.3.7 Tantangan Oversmoothing pada GNN Berlapis Dalam

Subbab 6.3.6 menunjukkan bahwa menambah jumlah lapisan propagasi memungkinkan setiap node menangkap konteks relasional yang semakin luas. Namun, intuisi "semakin dalam semakin baik" yang berlaku pada CNN (Bab 5) tidak sepenuhnya berlaku pada GNN. Ketika jumlah lapisan bertambah terlalu banyak, seluruh node dalam graf. Bahkan yang berada di bagian graf yang berjauhan secara bertahap mengonvergensi menuju representasi yang hampir identik, sebuah fenomena yang dikenal sebagai oversmoothing.

Secara intuitif, oversmoothing terjadi karena setiap lapisan propagasi pada dasarnya melakukan operasi pemulusan (smoothing) yang merata-ratakan representasi suatu node dengan representasi tetangganya, sebagaimana terlihat pada formulasi GCN (Subbab 6.3.2). Setelah diterapkan berulang kali pada banyak lapisan, efek pemulusan ini terakumulasi hingga representasi akhir setiap node didominasi oleh rata-rata keseluruhan graf, bukan lagi mencerminkan karakteristik lokal node tersebut. Secara matematis, dapat ditunjukkan bahwa representasi node $H^{(l)}$ pada GCN konvergen menuju subruang tertentu yang ditentukan oleh struktur graf itu sendiri (bukan lagi oleh fitur awal X) ketika $l \rightarrow \infty$, sehingga daya diskriminatif (discriminative power) antarnode menghilang.

Implikasi oversmoothing sangat relevan bagi konteks keamanan siber yang dibahas buku ini: jika model GNN yang digunakan pada studi kasus Bab 11 dirancang dengan lapisan yang terlalu dalam, representasi host yang benar-benar mencurigakan dapat menjadi tidak terbedakan dari representasi host normal di sekitarnya, justru mengaburkan sinyal anomali yang seharusnya ditangkap. Beberapa teknik mitigasi yang umum diterapkan dalam literatur meliputi:

1. **Jumping Knowledge (JK) Networks**. Alih-alih hanya menggunakan representasi dari lapisan terakhir, arsitektur ini menggabungkan (melalui concatenation, max-pooling, atau LSTM) representasi node dari seluruh lapisan perantara, sehingga model dapat secara adaptif memilih rentang informasi (range) yang paling sesuai untuk setiap node — beberapa node mungkin membutuhkan konteks lokal (lapisan dangkal), sementara node lain membutuhkan konteks yang lebih luas (lapisan dalam) (Xu et al., 2018).
2. **DropEdge**. Teknik regularisasi yang secara acak menghapus sebagian kecil edge dari graf pada setiap iterasi pelatihan, mengurangi kecepatan konvergensi menuju representasi yang seragam sekaligus berfungsi sebagai bentuk augmentasi data yang memperkuat generalisasi model, mirip prinsip dropout pada neural network konvensional (Bab 5) (Rong et al., 2020).
3. **Residual/Skip Connections**. Menambahkan koneksi langsung dari representasi awal atau lapisan sebelumnya ke lapisan saat ini, sehingga informasi asli node tetap terjaga meskipun melewati banyak lapisan propagasi, meminjam prinsip yang serupa dengan arsitektur residual network pada domain computer vision.
4. **Pembatasan kedalaman arsitektur (shallow GNN)**. Pendekatan paling sederhana namun tetap efektif: membatasi jumlah lapisan pada rentang 2–4 lapisan, yang pada praktiknya cukup untuk

menangkap pola pergerakan lateral berjarak beberapa hop tanpa mengorbankan daya diskriminatif, sejalan dengan konfigurasi yang telah diilustrasikan pada Pseudocode 6.2.

Kesadaran terhadap tantangan oversmoothing ini penting bagi pembaca yang hendak mengembangkan atau menyesuaikan arsitektur GNN pada Bab 11 maupun konteks lain: kedalaman arsitektur bukanlah hiperparameter yang dapat dinaikkan secara sembarangan demi menangkap konteks yang lebih luas, melainkan perlu diseimbangkan secara cermat dengan risiko hilangnya daya diskriminatif antarnode.

6.4 Prinsip Kerja: Representasi Graf dan Message Passing

Gambar 6.1 mengilustrasikan dua aspek fundamental GNN: (a) bagaimana data jaringan direpresentasikan sebagai graf, dengan jalur pergerakan lateral penyerang yang dapat ditelusuri melalui relasi antarnode; dan (b) prinsip message passing, mekanisme inti di balik seluruh varian GNN, di mana setiap node memperbarui representasinya dengan mengagregasi informasi dari node-node tetangganya secara berulang pada beberapa lapisan (multi-layer), sehingga secara bertahap menangkap konteks relasional yang semakin luas dan kompleks.

Gambar 6.1 Representasi Graf Jaringan dan Prinsip Kerja Graph Neural Network

Sumber: disintesis oleh penulis dari ACM TKDD (2023, dikutip dalam arXiv:2512.12112); Mitra et al. (2024); AboulEla & Kashef (2025).

Gambar 6.1 Representasi Graf Jaringan dan Prinsip Kerja Graph Neural Network (disintesis oleh penulis)

Tabel 6.1 merangkum tiga varian GNN yang paling banyak digunakan dalam penelitian keamanan siber, beserta mekanisme agregasi dan keunggulan masing-masing.

Tabel 6.1 Varian Graph Neural Network untuk Keamanan Siber

Varian GNN	Mekanisme Agregasi	Keunggulan	Contoh Penerapan Keamanan Siber
GCN (Graph Convolutional Network)	Rata-rata terbobot dari seluruh tetangga langsung	Efisien secara komputasi; baik untuk graf besar dan homogen	Deteksi pola transaksi mencurigakan pada jaringan bitcoin (Asiri & Somasundaram, 2025)
GAT (Graph Attention Network)	Bobot perhatian (attention) berbeda untuk tiap tetangga	Menonjolkan entitas paling relevan; lebih tahan terhadap noise relasi	Prioritisasi node berisiko tinggi dalam jaringan kompleks (arXiv, 2025)
GraphSAGE	Sampling tetangga + fungsi agregasi yang dapat digeneralisasi	Mampu melakukan inferensi pada node baru yang belum pernah dilihat saat pelatihan (inductive learning)	Deteksi entitas baru yang mencurigakan pada jaringan yang terus bertumbuh (Yan et al., 2023)

6.5 Implementasi: Pseudocode Pipeline GNN untuk Deteksi Pola Mencurigakan

Bagian ini menyajikan pseudocode implementasi menyeluruh yang menerapkan formulasi matematis pada Subbab 6.3, merepresentasikan

bagaimana pipeline GCN dibangun secara praktis untuk mendeteksi entitas mencurigakan dalam graf jaringan. Pendekatan yang mendasari studi kasus deteksi fraud Bitcoin pada Subbab 6.6 maupun studi kasus implementasi utama pada Bab 11.

Tahap 1: Konstruksi Graf dari Data Mentah

Pseudocode 6.1 — Konstruksi Graf dari Log Trafik Jaringan

```
INPUT: log_trafik (catatan koneksi antarentitas: sumber,
tujuan, atribut)
OUTPUT: A (matriks ketetanggaan), X (matriks fitur node)

FUNGSI bangun_graf(log_trafik):
    daftar_node = HimpunanUnik(
        log_trafik.sumber + log_trafik.tujuan)
    n = len(daftar_node)

    // Inisialisasi matriks ketetanggaan kosong berukuran n
x n
    A = MatriksNol(n, n)
    UNTUK setiap koneksi DALAM log_trafik:
        i = IndeksNode(koneksi.sumber, daftar_node)
        j = IndeksNode(koneksi.tujuan, daftar_node)
        A[i][j] = 1
        A[j][i] = 1     // graf tidak berarah (undirected)

    // Ekstraksi fitur setiap node (Subbab 6.3.1)
    X = MatriksKosong(n, dim_fitur)
    UNTUK setiap node DALAM daftar_node:
        X[IndeksNode(node)] = EkstrakFiturNode(node,
log_trafik)
        // contoh fitur: jumlah koneksi keluar/masuk, volume
data,
        // ragam protokol, riwayat anomali sebelumnya

    KEMBALIKAN A, X, daftar_node
```

Tahap 2: Pelatihan Model GCN

Pseudocode 6.2 — Pelatihan Graph Convolutional Network

```
INPUT: A, X, y_label (label node yang diketahui:
normal/mencurigakan)
OUTPUT: model_gcn (model GNN terlatih)

FUNGSI latih_gcn(A, X, y_label):
    // 1. Normalisasi matriks ketetanggaan (Subbab 6.3.2)
    A_tilde = A + MatriksIdentitas(len(A))      // tambah
self-loop
    D_tilde = MatriksDerajat(A_tilde)
    A_norm = InversSqrt(D_tilde) @ A_tilde @
InversSqrt(D_tilde)

    // 2. Definisikan arsitektur GCN dua lapis
    model_gcn = GCNModel(
        lapisan_1=GCNLayer(dim_input=X.shape(Asiri &
Somasundaram, 2025), dim_output=32),
        aktivasi=ReLU,
        lapisan_2=GCNLayer(dim_input=32,
dim_output=jumlah_kelas),
        aktivasi_akhir=Softmax
    )

    // 3. Latih dengan backpropagation (hanya pada node yang
berlabel)
    UNTUK epoch DALAM range(1, 200):
        H1 = model_gcn.lapisan_1.forward(A_norm, X)      //
H⁽¹⁾
        H2 = model_gcn.lapisan_2.forward(A_norm, H1)     //
H⁽²⁾
        loss = CrossEntropyLoss(H2[node_berlabel], y_label)
        gradien = Backpropagation(loss)
        model_gcn.perbarui_bobot(gradien,
learning_rate=0.01)
```

```
    KEMBALIKAN model_gcn
```

Tahap 3: Deteksi dan Skoring Node Mencurigakan

Pseudocode 6.3 — Inferensi dan Identifikasi Node Berisiko Tinggi

```
INPUT: model_gcn, A_norm, X (termasuk node baru yang belum
berlabel)
OUTPUT: daftar_node_berisiko (diurutkan berdasarkan skor
risiko)

FUNGSI deteksi_node_mencurigakan(model_gcn, A_norm, X,
daftar_node):
    probabilitas = model_gcn.forward(A_norm, X)   //
probabilitas per kelas

    hasil = []
    UNTUK setiap i DALAM range(len(daftar_node)):
        skor_mencurigakan =
probabilitas[i][KELAS_MENCURIGAKAN]
        JIKA skor_mencurigakan > ambang_batas:
            // Telusuri tetangga langsung sebagai konteks
investigasi
            // (mendukung penjelasan mirip prinsip XAI pada
Bab 7)
            tetangga_terkait = TelusuriTetangga(A_norm, i)
            hasil.tambahkan({
                node: daftar_node[i],
                skor: skor_mencurigakan,
                konteks_relasi: tetangga_terkait
            })

    daftar_node_berisiko = UrutkanMenurun(hasil,
kunci='skor')
```

```
    KEMBALIKAN daftar_node_berisiko
```

Pseudocode 6.1–6.3 menggambarkan siklus lengkap yang telah dibahas secara konseptual pada Gambar 6.1: konstruksi graf dari data mentah (Pseudocode 6.1) merepresentasikan proses transformasi yang juga akan muncul kembali pada arsitektur studi kasus Bab 11 (lihat Gambar 11.1, tahap 3); pelatihan model (Pseudocode 6.2) menerapkan langsung formulasi matematis GCN pada Subbab 6.3.2; dan deteksi node mencurigakan (Pseudocode 6.3) menunjukkan bagaimana keluaran model GNN yang pada dasarnya berupa probabilitas numerik diterjemahkan menjadi informasi yang dapat ditindaklanjuti analis keamanan, lengkap dengan konteks relasi antarentitas yang menjadi keunggulan utama pendekatan graph-based dibandingkan model konvensional yang telah dibahas pada Bab 4.

6.6 Studi Kasus: Deteksi Pola Kecurangan pada Transaksi Bitcoin Menggunakan Graph Convolutional Network

Konteks dan Permasalahan

Sebuah penelitian yang dipublikasikan pada Scientific Reports tahun 2025 mengangkat permasalahan deteksi kecurangan (fraud) pada jaringan transaksi mata uang kripto Bitcoin. Karakteristik unik dari domain ini adalah bahwa setiap transaksi secara inheren membentuk struktur graf alamat dompet digital sebagai node, dan transaksi sebagai edge yang menghubungkannya. Sehingga pola pencucian uang atau transaksi mencurigakan sering kali hanya dapat dikenali dari struktur relasi

antaralamat, bukan dari atribut transaksi individual semata (Asiri & Somasundaram, 2025).

Pendekatan

Peneliti menerapkan Graph Convolutional Network (GCN) untuk mempelajari representasi setiap alamat dompet berdasarkan pola transaksinya dengan alamat-alamat lain dalam jaringan. Model ini memungkinkan sistem mengidentifikasi kelompok alamat yang menunjukkan pola transaksi mencurigakan, misalnya rangkaian transfer bernilai kecil yang berulang dalam waktu singkat, sebuah pola umum pada skema pencucian uang digital. Meskipun setiap transaksi individual tampak tidak mencurigakan jika dianalisis secara terpisah.

Hasil dan Pembelajaran

Penelitian ini menunjukkan bahwa pendekatan berbasis graf secara signifikan mengungguli metode klasifikasi konvensional yang memperlakukan setiap transaksi secara independen, karena mampu memanfaatkan konteks relasional yang menjadi ciri khas skema kecurangan finansial yang terorganisir (Asiri & Somasundaram, 2025). Studi lain yang lebih baru turut menegaskan tren ini dengan mengintegrasikan pendekatan graph neural network bertopologi kuantum untuk mendeteksi pola kecurangan yang lebih kompleks, menunjukkan bahwa arah riset graph-based fraud detection terus berkembang pesat (arXiv, 2025).

Pembelajaran utama dari studi kasus ini bahwa pola ancaman/kecurangan sering kali hanya terlihat pada tingkat relasi antarentitas, bukan pada atribut individual menjadi prinsip inti yang mendasari rancangan studi kasus implementasi penulis pada Bab 11, di mana pendekatan serupa diterapkan untuk mendeteksi pola serangan siber pada jaringan komputer.

Catatan: Mengapa GNN Relevan untuk Studi Kasus Buku Ini

- *Serangan siber modern, sebagaimana dibahas pada Bab 1, semakin bersifat multi-tahap dan melibatkan pergerakan lateral antarentitas karakteristik yang secara alami dipetakan oleh struktur graf.*
- *Kajian survei tentang graph mining untuk keamanan siber menegaskan bahwa kemampuan GNN memodelkan relasi antarentitas menjadikannya pendekatan yang unggul untuk mendeteksi ancaman yang tersembunyi dalam struktur jaringan kompleks (Yan et al., 2023).*
- *GraphSAGE, dengan kemampuan inductive learning-nya, sangat relevan untuk lingkungan produksi di mana entitas jaringan baru (host, pengguna) terus bermunculan dan model perlu melakukan inferensi tanpa pelatihan ulang penuh.*
- *Studi kasus GNN pada bab ini baik deteksi fraud Bitcoin maupun prinsip pergerakan lateral menjadi fondasi konseptual langsung bagi Intelligent Cybersecurity Mitigation Recommendation System yang dibahas pada Bab 11.*

6.8 Rangkuman

GNN menangani keterbatasan mendasar pendekatan konvensional dengan memodelkan data keamanan siber sebagai graf entitas dan relasi, bukan baris data yang independen. Prinsip inti seluruh varian GNN adalah message passing: setiap node memperbarui representasinya dengan mengagregasi informasi dari node tetangga secara berlapis. GCN unggul dalam efisiensi pada graf besar, GAT unggul dalam menonjolkan relasi paling relevan melalui mekanisme attention, dan GraphSAGE unggul dalam kemampuan

inferensi pada entitas baru. Studi kasus deteksi fraud Bitcoin menggunakan GCN menunjukkan keunggulan nyata pendekatan graph-based dalam mengenali pola kecurangan terorganisir yang tidak terlihat pada level transaksi individual. Prinsip-prinsip GNN pada bab ini menjadi fondasi langsung bagi studi kasus implementasi sistem rekomendasi mitigasi berbasis GNN pada Bab 11.

Daftar Pustaka

Asiri, A., & Somasundaram, K. (2025). Graph Convolution Network for Fraud Detection in Bitcoin Transactions. Scientific Reports, 15(1), 11076.

arXiv. (2025). Quantum Topological Graph Neural Networks for Detecting Complex Fraud Patterns. arXiv:2512.03696.

Yan, B., Yang, C., Shi, C., Fang, Y., Li, Q., Ye, Y., & Du, J. (2023). Graph Mining for Cybersecurity: A Survey. ACM TKDD, 18(2).

Mitra, S., Chakraborty, T., Neupane, S., Piplai, A., & Mittal, S. (2024). Use of Graph Neural Networks in Aiding Defensive Cyber Operations. arXiv:2401.05680.

AboulEla, S. G., & Kashef, R. F. (2025). Leveraging Large Language Models, Graph Neural Networks, and Explainable AI for Next-Generation Network Intrusion Detection Systems. Journal of Intelligent Information Systems, 63, 1807–1835.

Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K., & Jegelka, S. (2018). Representation Learning on Graphs with Jumping Knowledge Networks. Proceedings of the 35th International Conference on Machine Learning (ICML).

Rong, Y., Huang, W., Xu, T., & Huang, J. (2020). DropEdge: Towards Deep Graph Convolutional Networks on Node Classification. International Conference on Learning Representations (ICLR).

BAB 7

Explainable AI dan Large Language Models dalam Konteks Keamanan Siber

7.1 Pendahuluan

Bab-bab sebelumnya telah menunjukkan bahwa model machine learning, deep learning, dan graph neural network mampu mencapai akurasi deteksi yang sangat tinggi. Namun, akurasi saja tidak cukup untuk penerapan operasional: analis keamanan perlu memahami mengapa sebuah keputusan diambil sebelum bertindak berdasarkan rekomendasi model, terutama untuk tindakan berdampak tinggi seperti mengisolasi server produksi. Bab ini membahas dua pendekatan yang saling melengkapi untuk menjembatani kesenjangan ini: explainable AI (XAI) dan large language model (LLM).

7.2 Mengapa Explainable AI Menjadi Kebutuhan, Bukan Pilihan

Model deep learning dan ensemble yang dibahas pada bab-bab sebelumnya secara umum bersifat black box keputusannya sulit ditelusuri kembali ke faktor-faktor yang mendasarinya. Kajian tinjauan literatur XAI untuk keamanan siber menunjukkan bahwa opasitas ini memunculkan tiga

tantangan utama: analis keamanan kesulitan memvalidasi dan mempercayai peringatan yang dihasilkan AI, organisasi kesulitan memenuhi kewajiban kepatuhan regulasi seperti GDPR yang menuntut akuntabilitas keputusan otomatis, dan kolaborasi antara manusia dan AI menjadi terhambat karena kurangnya dasar kepercayaan bersama (Frontiers in Computer Science, 2026).

Tabel 7.1 merangkum tiga teknik XAI yang paling banyak diadopsi dalam penelitian dan praktik keamanan siber terkini.

Tabel 7.1 Teknik Explainable AI (XAI) yang Umum Digunakan dalam Keamanan Siber

Teknik XAI	Prinsip Kerja	Kegunaan dalam Keamanan Siber
SHAP (Shapley Additive Explanations)	Mengukur kontribusi setiap fitur terhadap prediksi berdasarkan teori permainan kooperatif	Menjelaskan mengapa trafik tertentu diklasifikasikan sebagai serangan, fitur mana yang paling berpengaruh
LIME (Local Interpretable Model-agnostic Explanations)	Membangun model sederhana yang mudah diinterpretasi di sekitar satu prediksi spesifik	Memberikan penjelasan lokal untuk kasus tunggal yang mudah dipahami analis
Attention Mechanism	Menunjukkan bobot perhatian model terhadap bagian input tertentu	Menyoroti byte, log, atau segmen sekuens yang paling memengaruhi keputusan model deep learning

7.3 Large Language Model sebagai Jembatan Komunikasi Analisis Ancaman

Selain XAI, large language model menawarkan pendekatan pelengkap: alih-alih hanya menjelaskan keputusan model lain, LLM dapat secara aktif memproses, meringkas, dan mengomunikasikan informasi ancaman dalam bahasa natural yang mudah dipahami. Tabel 7.2 merangkum sejumlah kapabilitas LLM yang relevan bagi operasi keamanan siber.

Tabel 7.2 Kapabilitas Large Language Model dalam Mendukung Operasi Keamanan Siber

Kapabilitas LLM	Contoh Penerapan	Manfaat Operasional
Interpretasi Alert dalam Bahasa Natural	ChatNIDS — menerjemahkan alert NIDS teknis menjadi penjelasan intuitif	Membantu analis pemula memahami dan menindaklanjuti peringatan tanpa keahlian mendalam (arXiv, 2025a)
Agen Investigasi Otomatis	Agen LLM melakukan triase awal alert sebelum eskalasi ke analis manusia	Mengurangi beban kerja analis pada volume alert yang tinggi (arXiv, 2025b)
Analisis Threat Intelligence	Peringkasan dan korelasi laporan ancaman dari berbagai sumber	Mempercepat sintesis informasi ancaman yang tersebar di banyak dokumen
Kolaborasi dengan NIDS	LLM dikombinasikan dengan model neural network (mis. LSTM classifier) untuk analisis konteks	Memperkuat pemahaman kontekstual atas serangan bertahap (Bab 2)

7.4 Studi Kasus: CORTEX — Kolaborasi Agen LLM untuk Triase Alert Berisiko Tinggi

Konteks dan Permasalahan

Sebuah studi arXiv tahun 2025 mengangkat permasalahan klasik pada security operations center (SOC): volume alert yang sangat tinggi menyebabkan fenomena alert fatigue, di mana penelitian sebelumnya bahkan mencatat bahwa hingga 99% alert yang dihasilkan sistem keamanan ternyata merupakan false positive, membuat analis kewalahan dan berisiko melewatkan ancaman yang sesungguhnya kritis (arXiv, 2025b).

Pendekatan

Peneliti mengembangkan CORTEX, sebuah kerangka kerja yang menggunakan beberapa agen LLM yang berkolaborasi untuk melakukan triase alert berisiko tinggi secara bertahap — satu agen berfokus pada pengumpulan konteks tambahan, agen lain melakukan penalaran terhadap tingkat keparahan, dan agen ketiga menyusun rekomendasi tindakan — meniru pola kerja tim SOC manusia namun pada skala dan kecepatan yang jauh lebih tinggi.

Hasil dan Pembelajaran

Pendekatan multi-agen ini menunjukkan bahwa LLM tidak harus digunakan sebagai kotak hitam tunggal yang membuat keputusan akhir, melainkan dapat disusun sebagai sistem kolaboratif dengan pembagian peran yang meniru praktik terbaik investigasi keamanan siber manusia, sekaligus tetap mempertahankan jejak penalaran yang dapat ditelusuri (traceable). Sebuah bentuk transparansi yang melengkapi teknik XAI klasik seperti SHAP dan LIME (arXiv, 2025b).

Studi lain yang lebih luas, sebuah survei sistematis terhadap 153 studi peer-review tentang AI-augmented SOC, menegaskan bahwa meskipun XAI menawarkan transparansi, adopsinya oleh incident responder di lapangan masih terbatas karena kompleksitas antarmuka penjelasan yang dihasilkan

menunjukkan bahwa tantangan XAI bukan hanya teknis, melainkan juga soal desain interaksi manusia-AI yang efektif (ResearchGate / MDPI, 2025).

Catatan: Pentingnya Sinergi XAI dan LLM

- *XAI klasik (SHAP, LIME, attention) unggul dalam menjelaskan kontribusi fitur secara matematis presisi, namun keluarannya sering kali masih terlalu teknis untuk dipahami analis non-pakar (Frontiers in Computer Science, 2026).*
- *LLM unggul dalam mengomunikasikan hasil analisis dalam bahasa natural yang mudah dipahami, namun tanpa dasar penjelasan yang solid, LLM berisiko menghasilkan justifikasi yang terdengar meyakinkan namun tidak akurat (halusinasi) (ResearchGate / MDPI, 2025).*
- *Kombinasi keduanya — XAI untuk dasar penjelasan yang presisi, LLM untuk komunikasi yang mudah dipahami — menawarkan pendekatan yang saling melengkapi, sebagaimana ditunjukkan oleh kerangka kerja seperti ChatNIDS (arXiv, 2025a).*
- *Prinsip transparansi pada bab ini menjadi komponen penting dalam rancangan sistem rekomendasi mitigasi adaptif pada Bab 9 dan studi kasus implementasi pada Bab 11, di mana rekomendasi yang dihasilkan AI perlu disertai penjelasan yang dapat dipercaya praktisi keamanan.*

7.5 Rangkuman

Explainable AI (XAI) menjadi kebutuhan mendasar untuk mengatasi sifat black box model deep learning dan ensemble, memungkinkan analis memvalidasi dan mempercayai keputusan AI. SHAP, LIME, dan attention

mechanism merupakan tiga teknik XAI yang paling banyak diadopsi, masing-masing dengan pendekatan matematis yang berbeda namun tujuan yang sama: transparansi keputusan model. Large language model menawarkan kapabilitas pelengkap berupa komunikasi hasil analisis dalam bahasa natural, interpretasi alert, dan investigasi otomatis awal. Studi kasus CORTEX menunjukkan bahwa kolaborasi multi-agen LLM dapat meniru pola kerja tim SOC manusia sekaligus mempertahankan jejak penalaran yang dapat ditelusuri. Adopsi XAI di lapangan masih menghadapi tantangan desain interaksi manusia-AI, menunjukkan bahwa transparansi teknis saja tidak cukup tanpa antarmuka yang dapat dipahami penggunanya.

Daftar Pustaka

arXiv. (2025a). Network Intrusion Detection: Evolution from Conventional Approaches to LLM Collaboration and Emerging Risks. arXiv:2510.23313.

arXiv. (2025b). CORTEX: Collaborative LLM Agents for High-Stakes Alert Triage. arXiv:2510.00311.

Frontiers in Computer Science. (2026). Explainable AI: Enhancing Decision-Making in the Detection of Cyber Threats.

ResearchGate / MDPI. (2025). AI-Augmented SOC: A Survey of LLMs and Agents for Security Automation.

Issues in Information Systems. (2025). Bridging Advanced Threat Detection and Human Understanding Through Explainable Artificial Intelligence, Vol. 26, No. 3.

BAB 8

Sistem Deteksi Intrusi Berbasis AI (IDS/IPS Cerdas)

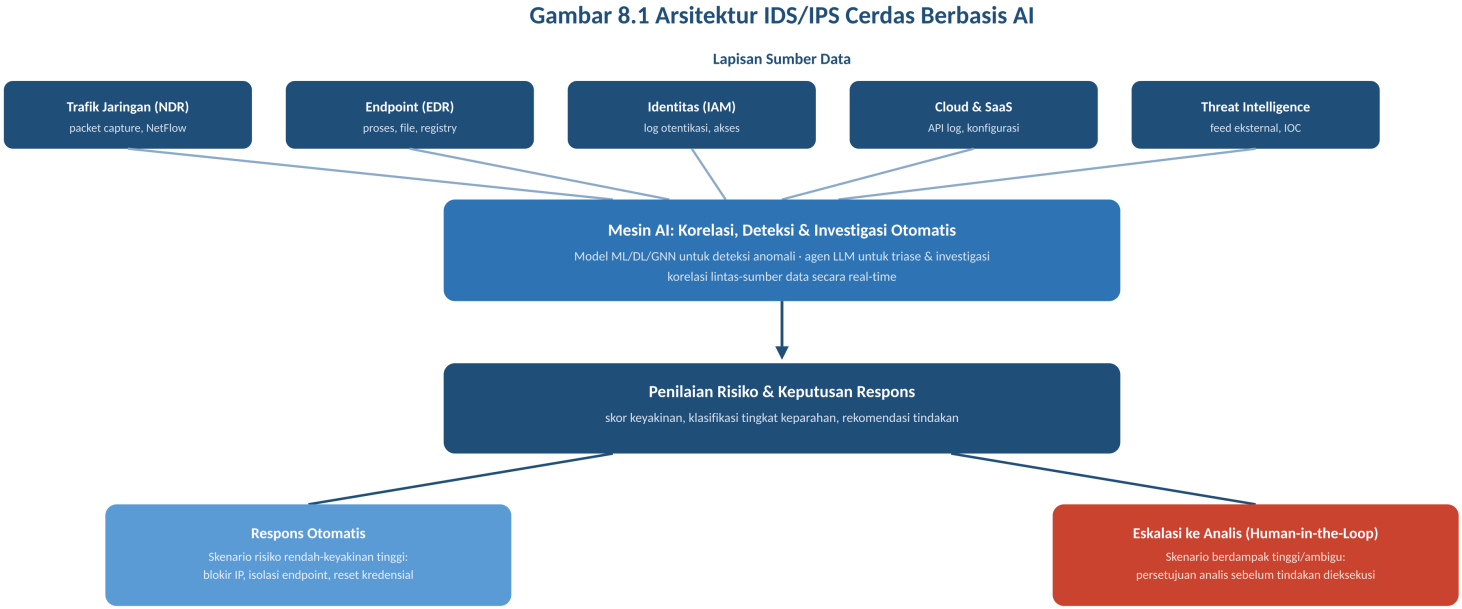
8.1 Pendahuluan

Bagian II buku ini telah membahas fondasi teknis kecerdasan buatan — machine learning, deep learning, graph neural network, serta explainable AI dan large language model. Bab ini membuka Bagian III dengan membahas bagaimana fondasi-fondasi tersebut diintegrasikan menjadi sistem operasional nyata: intrusion detection system (IDS) dan intrusion prevention system (IPS) cerdas yang mampu mendeteksi sekaligus merespons ancaman secara real-time.

8.2 Dari IDS/IPS Tradisional menuju SOC Berbasis AI

IDS tradisional, sebagaimana dibahas pada Bab 2, umumnya hanya memberi peringatan (alert) tanpa mengambil tindakan, sementara IPS menambahkan kemampuan pemblokiran otomatis berdasarkan aturan statis. Generasi terbaru sistem ini yang kini banyak disebut sebagai AI SOC platform mengintegrasikan model AI di setiap lapisan operasi: korelasi data lintas-sumber, investigasi otomatis, penilaian risiko, hingga eksekusi respons bertingkat.

Gambar 8.1 mengilustrasikan arsitektur umum IDS/IPS cerdas berbasis AI, mulai dari pengumpulan data lintas-sumber (jaringan, endpoint, identitas, cloud, threat intelligence), pemrosesan oleh mesin AI untuk korelasi dan investigasi, hingga keputusan respons yang dapat dieksekusi secara otomatis penuh atau dieskalasi kepada analis manusia tergantung tingkat risiko dan keyakinan model.



Sumber: disintesis oleh penulis dari Radiant Security (2026), Torq HyperSOC (2026), Exaforce SecOps Radar (2026), dan Vectra AI SOC Automation Guide (2026).

Gambar 8.1 Arsitektur IDS/IPS Cerdas Berbasis AI (disintesis oleh penulis)

8.3 Lanskap Platform AI SOC Terkini

Pasar platform SOC berbasis AI berkembang sangat pesat pada 2025–2026. Tabel 8.1 merangkum sejumlah platform terkemuka beserta kapabilitas dan metrik kinerja yang mereka laporkan.

Tabel 8.1 Perbandingan Platform IDS/IPS Cerdas Berbasis AI (2025–2026)

Platform	Kapabilitas Kunci	Metrik Kinerja yang Dilaporkan
CrowdStrike Charlotte AI	Triase deteksi otomatis, agentic response untuk investigasi & containment tingkat analis	Akurasi penilaian alert otomatis >98%, dilatih dari keputusan analis MDR elite (UnderDefense, 2026)
Torq HyperSOC	Orkestrasi multi-agen (Socrates) untuk pengayaan, manajemen kasus, verifikasi, remediasi	Integrasi tanpa kode ke SIEM/EDR/IAM; siklus kasus end-to-end terkoordinasi (Torq, 2026)
Conifers CognitiveSOC	Arsitektur mesh agentic AI untuk investigasi multi-tingkat	87% investigasi lebih cepat, throughput SOC 3x lipat, akurasi >99% (Conifers.ai, 2026)
ReliaQuest (otomasi umum industri)	Otomasi respons berbasis AI pada volume alert tinggi	Waktu respons di bawah 7 menit, dibandingkan 2,3 hari tanpa otomasi (Vectra AI, 2026)

8.4 Studi Kasus: Deteksi Ancaman Orang Dalam (Insider Threat) dan Otomasi Respons Insiden

Konteks dan Permasalahan

Salah satu tantangan tersulit bagi SOC tradisional adalah mendeteksi ancaman orang dalam (insider threat) karyawan atau pihak berwenang yang menyalahgunakan akses sahnya untuk mengeksfiltrasi data sensitif. Ancaman ini sulit dideteksi oleh signature-based detection karena pelaku menggunakan kredensial yang sah, dan sulit dibedakan dari perubahan pola kerja yang wajar tanpa menghasilkan banyak false positive (Radiant Security, 2026).

Pendekatan

Platform AI SOC modern mengatasi tantangan ini dengan membangun profil perilaku historis setiap karyawan dan kelompok sejawatnya (peer group), kemudian menandai penyimpangan signifikan, seperti pola akses berkas yang tidak wajar, aktivitas di luar jam kerja, atau volume transfer data yang jauh melampaui kebiasaan sebagai potensi ancaman. Ketika insiden terkonfirmasi, sistem melakukan analisis dampak komprehensif untuk mengidentifikasi pengguna, perangkat, dan aplikasi yang terdampak, lalu menghasilkan rencana respons yang dapat dieksekusi secara manual, semi-otomatis, atau sepenuhnya otomatis (Radiant Security, 2026).

Hasil dan Pembelajaran

Kemampuan sistem untuk membedakan antara perubahan pola kerja yang sah dan aktivitas yang benar-benar mencurigakan menjadi kunci keberhasilan pendekatan ini mengurangi tuduhan yang tidak berdasar (false accusation) sekaligus tetap menangkap ancaman yang genuine (Radiant Security, 2026). Pada dimensi kecepatan respons, data industri menunjukkan bahwa otomasi berbasis AI mampu mengeksekusi tindakan containment isolasi endpoint, pemblokiran IP berbahaya, reset kredensial dalam hitungan detik setelah deteksi, jauh lebih cepat dibandingkan proses manual yang dapat memakan waktu berhari-hari (Radiant Security, 2026).

Survei industri tahun 2026 turut mencatat bahwa organisasi dengan otomasi AI yang matang mencapai waktu respons rata-rata di bawah 7 menit, dibandingkan 2,3 hari pada organisasi tanpa otomasi sebuah perbedaan yang secara langsung menentukan skala kerusakan yang dapat dicegah (Vectra AI, 2026).

Catatan: Human-in-the-Loop Tetap Esensial

- *Meskipun otomasi penuh menawarkan kecepatan, praktik terbaik industri tetap mempertahankan model human-in-the-loop untuk tindakan berdampak tinggi seperti menonaktifkan akun pengguna, sementara tindakan berisiko rendah dan berkeyakinan tinggi (seperti memblokir IP yang telah dikenal berbahaya) dapat dieksekusi sepenuhnya otomatis.*
- *Gartner memprediksi bahwa 40% aplikasi enterprise akan memiliki agen AI tugas-spesifik pada akhir 2026, meningkat dari kurang dari 5% pada 2025 menandakan bahwa integrasi AI ke dalam alur kerja keamanan siber akan semakin menjadi standar, bukan pengecualian (Vectra AI, 2026).*
- *Regulasi seperti Cyber Incident Reporting for Critical Infrastructure Act (CIRCIA) di Amerika Serikat, yang mewajibkan pelaporan insiden signifikan dalam 72 jam, turut mendorong kebutuhan mendesak akan sistem deteksi dan pelaporan otomatis (Vectra AI, 2026).*

8.5 Rangkuman

IDS/IPS cerdas berbasis AI mengintegrasikan model ML/DL/GNN dan agen LLM ke seluruh lapisan operasi: pengumpulan data, korelasi, investigasi, penilaian risiko, hingga eksekusi respons. Lanskap platform AI SOC 2025–

2026 menunjukkan kematangan operasional yang tinggi, dengan metrik kinerja seperti akurasi triase >98% dan investigasi 87% lebih cepat pada platform terkemuka. Studi kasus deteksi ancaman orang dalam menunjukkan bagaimana AI dapat membedakan perilaku wajar dari yang mencurigakan, mengurangi false accusation sekaligus mempercepat containment insiden. Model human-in-the-loop tetap menjadi praktik terbaik untuk tindakan berdampak tinggi, sejalan dengan pembahasan transparansi dan kepercayaan pada Bab 7.

Daftar Pustaka

UnderDefense. (2026). 9 Best AI SOC Providers in 2026: A Complete Vendor Comparison.

Torq. (2026). 10 Best SOC Tools (2025): Legacy vs Modern Automation.

Conifers.ai. (2026). Top 10 AI SOC Agents, Platforms and Solutions in 2026.

Vectra AI. (2026). SOC Automation: Benefits, Tools, Use Cases & AI Trends (2026 Guide).

Radiant Security. (2026). Real-World Use Cases of AI-Powered SOC.

BAB 9

Sistem Rekomendasi Mitigasi Serangan Siber yang Adaptif

9.1 Pendahuluan

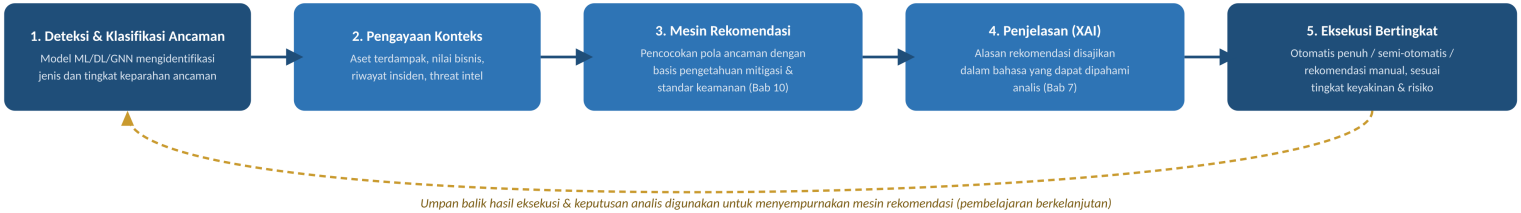
Bab 8 telah membahas bagaimana AI SOC platform mendeteksi dan merespons ancaman secara real-time. Namun, deteksi yang akurat baru menjawab separuh persoalan: organisasi juga membutuhkan panduan mengenai tindakan mitigasi apa yang paling tepat untuk setiap jenis ancaman yang terdeteksi. Bab ini membahas secara khusus bagaimana sistem rekomendasi mitigasi adaptif dirancang sebuah topik yang menjadi jembatan langsung menuju studi kasus implementasi penulis pada Bab 11.

9.2 Dari Deteksi menuju Rekomendasi: Pergeseran Paradigma

Banyak sistem keamanan konvensional berhenti pada tahap deteksi dan memberi peringatan generik, meninggalkan analis untuk secara manual memutuskan tindakan yang tepat berdasarkan pengalaman dan pengetahuan mereka sendiri. Sistem rekomendasi mitigasi adaptif berupaya melangkah lebih jauh: tidak hanya mendeteksi bahwa "terjadi anomali", melainkan juga menyarankan "apa yang sebaiknya dilakukan", disesuaikan dengan konteks spesifik ancaman, aset yang terdampak, dan tingkat risiko organisasi.

Gambar 9.1 menggambarkan arsitektur umum sistem rekomendasi mitigasi adaptif, terdiri dari lima tahap utama yang saling terhubung melalui mekanisme umpan balik berkelanjutan.

Gambar 9.1 Arsitektur Sistem Rekomendasi Mitigasi Serangan Siber yang Adaptif



Sumber: dirancang oleh penulis, disintesis dari arsitektur Radiant Security (2026) dan prinsip rekomendasi mitigasi kontekstual yang diusulkan pada penelitian penulis (Bab 11).

Gambar 9.1 Arsitektur Sistem Rekomendasi Mitigasi Serangan Siber yang Adaptif (dirancang oleh penulis)

Kelima tahap tersebut adalah: (1) deteksi dan klasifikasi ancaman menggunakan model ML/DL/GNN yang telah dibahas pada Bagian II; (2) pengayaan konteks dengan informasi aset, riwayat insiden, dan threat intelligence; (3) mesin rekomendasi yang mencocokkan pola ancaman dengan basis pengetahuan mitigasi dan standar keamanan (dibahas mendalam pada Bab 10); (4) penyajian penjelasan menggunakan prinsip XAI dari Bab 7 agar rekomendasi dapat dipahami dan dipercaya; dan (5) eksekusi bertingkat sesuai tingkat keyakinan dan risiko, sebagaimana dibahas pada Bab 8.

9.3 Kategori Rekomendasi Mitigasi berdasarkan Jenis Ancaman

Tabel 9.1 menyajikan contoh pemetaan antara kategori ancaman dan jenis rekomendasi mitigasi yang umum diterapkan, beserta tingkat otomasi yang lazim digunakan pada praktik industri.

Tabel 9.1 Kategori Ancaman dan Contoh Rekomendasi Mitigasi Adaptif

Kategori Ancaman	Contoh Rekomendasi Mitigasi Adaptif	Tingkat Otomasi Umum
Malware/Ransomware terkonfirmasi	Isolasi endpoint, hentikan proses mencurigakan, blokir hash file	Otomatis penuh (keyakinan tinggi)
Anomali akses/kredensial	Paksa reset kata sandi, aktifkan MFA tambahan, batasi sesi	Semi-otomatis (perlu verifikasi)
Pergerakan lateral terdeteksi (GNN)	Segmentasi jaringan sementara, pantau ketat node terkait	Rekomendasi + persetujuan analis

Anomali orang dalam (insider)	Tinjauan akses, notifikasi ke manajer/HR, pembekuan sementara akses berkas sensitif	Manual dengan dukungan konteks AI
Indikasi kerentanan belum ditambal	Prioritisasi patch berdasarkan eksploitabilitas, kontrol kompensasi sementara	Rekomendasi terprioritisasi

9.4 Studi Kasus: Rekomendasi Respons Kontekstual pada Platform AI SOC

Konteks dan Permasalahan

Melanjutkan studi kasus deteksi ancaman orang dalam pada Bab 8, aspek yang secara khusus relevan bagi bab ini adalah bagaimana platform AI SOC modern menghasilkan rencana respons yang disesuaikan (customized response plan) untuk setiap insiden yang terkonfirmasi, alih-alih menerapkan satu playbook generik untuk semua jenis insiden (Radiant Security, 2026).

Pendekatan

Ketika insiden terkonfirmasi, sistem melakukan analisis dampak komprehensif mengidentifikasi pengguna, perangkat, dan aplikasi yang terdampak sebelum menyusun rencana respons yang mempertimbangkan konteks spesifik tersebut. Rencana ini kemudian dapat dieksekusi secara manual (analis menjalankan setiap langkah), semi-otomatis (sistem menjalankan sebagian, meminta persetujuan untuk langkah berisiko tinggi), atau sepenuhnya otomatis (untuk skenario dengan keyakinan sangat tinggi dan risiko rendah) (Radiant Security, 2026). Sebuah

keunggulan operasional penting yang dicatat pada pendekatan ini adalah kemampuan sistem beradaptasi secara otonom terhadap perubahan lingkungan organisasi berdasarkan data telemetri real-time, tanpa memerlukan pembaruan manual playbook secara terus-menerus mengatasi salah satu beban operasional terbesar pada pendekatan SOAR generasi sebelumnya (Radiant Security, 2026).

Hasil dan Pembelajaran

Studi kasus ini menggarisbawahi prinsip inti dari sistem rekomendasi mitigasi adaptif: kemampuan menyesuaikan rekomendasi terhadap konteks spesifik (bukan playbook generik satu ukuran untuk semua), serta kemampuan belajar dan beradaptasi secara berkelanjutan seiring perubahan pola ancaman dan lingkungan organisasi dua prinsip yang secara langsung diadopsi dalam rancangan Intelligent Cybersecurity Mitigation Recommendation System pada Bab 11.

Catatan: Prinsip Perancangan Sistem Rekomendasi yang Efektif

- *Rekomendasi harus kontekstual mempertimbangkan nilai bisnis aset yang terdampak, bukan hanya jenis ancaman secara teknis semata.*
- *Tingkat otomasi harus proporsional terhadap tingkat keyakinan model dan dampak potensial tindakan, sejalan dengan prinsip human-in-the-loop yang dibahas pada Bab 8.*
- *Setiap rekomendasi idealnya disertai penjelasan (Bab 7) agar analis dapat memvalidasi kesesuaiannya sebelum eksekusi, khususnya untuk tindakan yang tidak dapat dibatalkan.*
- *Mekanisme umpan balik dari hasil eksekusi dan keputusan analis harus digunakan untuk menyempurnakan mesin rekomendasi secara*

berkelanjutan, menjadikan sistem semakin adaptif dari waktu ke waktu.

9.5 Rangkuman

Sistem rekomendasi mitigasi adaptif melangkah lebih jauh dari deteksi semata, dengan menyarankan tindakan mitigasi yang disesuaikan konteks ancaman dan aset terdampak. Arsitektur sistem terdiri dari lima tahap: deteksi, pengayaan konteks, mesin rekomendasi, penjelasan (XAI), dan eksekusi bertingkat dengan umpan balik berkelanjutan. Tingkat otomasi rekomendasi bervariasi tergantung tingkat keyakinan dan risiko, dari otomatis penuh hingga memerlukan persetujuan analis. Studi kasus platform AI SOC modern menunjukkan bahwa rencana respons yang disesuaikan konteks dan kemampuan adaptasi otonom terhadap perubahan lingkungan mengatasi keterbatasan pendekatan playbook generik generasi sebelumnya.

Daftar Pustaka

Radiant Security. (2026). Real-World Use Cases of AI-Powered SOC.

Exaforce. (2026). Best AI-Powered SOC Platforms 2025: A Comprehensive Guide for Security Leaders.

Vectra AI. (2026). SOC Automation: Benefits, Tools, Use Cases & AI Trends (2026 Guide).

BAB 10

Integrasi Domain Knowledge dan Standar Keamanan Internasional (NIST, ISO 27001)

10.1 Pendahuluan

Sistem rekomendasi mitigasi adaptif yang dibahas pada Bab 9 hanya akan bernilai praktis jika rekomendasi yang dihasilkan selaras dengan standar dan kerangka kerja keamanan yang telah diakui secara internasional. Bab ini membahas bagaimana keluaran sistem berbasis AI dapat diintegrasikan dengan dua kerangka kerja yang paling banyak diadopsi secara global: NIST Cybersecurity Framework (CSF) dan ISO/IEC 27001, sehingga rekomendasi yang dihasilkan tidak hanya akurat secara teknis, tetapi juga relevan secara tata kelola dan kepatuhan (compliance).

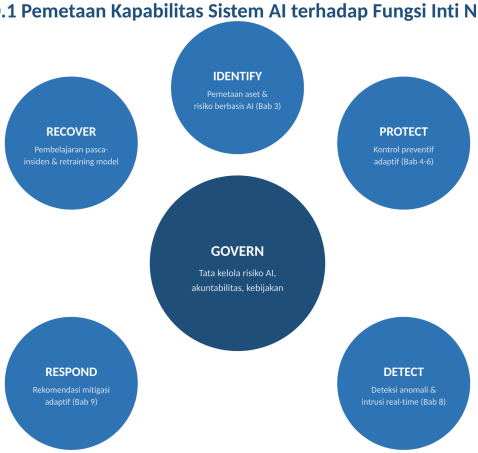
10.2 NIST Cybersecurity Framework 2.0: Fungsi Inti dan Peran AI

NIST CSF mengalami pembaruan signifikan pada versi 2.0 yang dirilis Februari 2024 revisi besar pertama sejak kerangka kerja ini pertama kali diperkenalkan pada 2014. Perubahan paling mendasar adalah penambahan fungsi keenam, Govern, yang ditempatkan sebagai fungsi lintas (cross-cutting) yang menaungi kelima fungsi operasional lainnya: Identify, Protect,

Detect, Respond, dan Recover. Penambahan fungsi Govern mencerminkan pergeseran pandangan bahwa keamanan siber merupakan risiko tingkat perusahaan yang harus dipimpin dan didanai di level eksekutif, bukan sekadar tanggung jawab tim teknis (SaltyCloud / Isora GRC, 2026; Expert Insights, 2026).

Gambar 10.1 memetakan bagaimana kapabilitas sistem berbasis AI yang dibahas di sepanjang buku ini berkontribusi pada masing-masing fungsi inti NIST CSF 2.0.

Gambar 10.1 Pemetaan Kapabilitas Sistem AI terhadap Fungsi Inti NIST CSF 2.0



Govern menjadi fungsi lintas seluruh fungsi operasional lainnya, sejalan dengan struktur NIST CSF 2.0 (2024)

Sumber: disintesis oleh penulis dari NIST Cybersecurity Framework 2.0 (2024) dan draf NIST Cyber AI Profile (2026).

Gambar 10.1 Pemetaan Kapabilitas Sistem AI terhadap Fungsi Inti NIST CSF 2.0 (disintesis oleh penulis)

Tabel 10.1 merinci deskripsi masing-masing fungsi beserta kontribusi konkret sistem berbasis AI terhadapnya.

Tabel 10.1 Fungsi Inti NIST CSF 2.0 dan Kontribusi Sistem Berbasis AI

Fungsi NIST CSF 2.0	Deskripsi Singkat	Kontribusi Sistem Berbasis AI
Govern	Tata kelola risiko siber, akuntabilitas, kebijakan tingkat eksekutif	Menyediakan jejak audit (audit trail) keputusan AI serta laporan risiko yang dapat ditelusuri untuk kebutuhan tata kelola (SaltyCloud / Isora GRC, 2026; Expert Insights, 2026)
Identify	Pemetaan aset, sistem, dan risiko yang perlu dilindungi	Model AI memetakan aset dan mengidentifikasi ketergantungan melalui representasi graf entitas (Bab 6)
Protect	Penerapan kontrol pencegahan seperti akses, enkripsi, pelatihan kesadaran	Model prediktif membantu memprioritaskan kontrol berdasarkan tingkat eksploitabilitas kerentanan (Bab 4)
Detect	Deteksi dini insiden dan aktivitas anomali	Inti dari seluruh model ML/DL/GNN yang dibahas pada Bagian II buku ini
Respond	Tindakan penanganan insiden yang telah terdeteksi	Sistem rekomendasi mitigasi adaptif menghasilkan tindakan respons kontekstual (Bab 9)
Recover	Pemulihan layanan dan pembelajaran pascainsiden	Data insiden digunakan untuk melatih ulang (retraining) model secara berkelanjutan

10.3 Perkembangan Terbaru: NIST Cyber AI Profile

Merespons pesatnya adopsi AI dalam keamanan siber, NIST tengah mengembangkan Cyber AI Profile sebuah perluasan sukarela dari CSF 2.0

yang secara khusus dirancang untuk mengatasi risiko dan peluang unik yang dihadirkan AI, dengan mengacu pula pada NIST AI Risk Management Framework (AI RMF). Draf profil ini disusun mengelilingi tiga area fokus: Secure (mengamankan sistem AI itu sendiri), Defend (menyelenggarakan pertahanan siber berbasis AI), dan Thwart (menggagalkan serangan siber berbasis AI menggunakan AI), yang seluruhnya tetap dipetakan terhadap keenam fungsi inti CSF 2.0 (KPMG, 2026).

Salah satu penekanan penting dalam draf ini adalah pada fungsi Respond, di mana NIST merekomendasikan organisasi untuk mempertahankan log, input, output, dan rantai keputusan sistem AI secara utuh guna memastikan ketertelusuran (provenance) dan mendukung penyempurnaan tindakan respons berbasis AI di masa depan sebuah rekomendasi yang secara langsung relevan dengan prinsip explainable AI yang telah dibahas pada Bab 7 (KPMG, 2026).

10.4 ISO/IEC 27001:2022: Integrasi dengan Kontrol Teknologi

Sementara NIST CSF menawarkan kerangka kerja manajemen risiko tingkat tinggi, ISO/IEC 27001 menyediakan spesifikasi kontrol yang lebih rinci dan dapat disertifikasi. Revisi 2022 mengorganisasikan seluruh kontrol Annex A ke dalam empat kategori. Tabel 10.2 merangkum keempat kategori tersebut beserta relevansinya terhadap sistem berbasis AI.

Tabel 10.2 Kategori Kontrol ISO/IEC 27001:2022 Annex A dan Relevansi terhadap Sistem AI

Kategori Kontrol ISO/IEC 27001:2022 Annex A	Jumlah Kontrol	Contoh Relevansi dengan Sistem AI

A.5 Kontrol Organisasi	37 kontrol	Kebijakan keamanan informasi, manajemen risiko pihak ketiga (relevan untuk model AI pihak ketiga/API)
A.6 Kontrol Manusia	8 kontrol	Pelatihan kesadaran keamanan, termasuk literasi terhadap rekomendasi yang dihasilkan AI
A.7 Kontrol Fisik	14 kontrol	Keamanan infrastruktur yang menjalankan model AI dan penyimpanan data pelatihan
A.8 Kontrol Teknologi	34 kontrol	Manajemen kerentanan, deteksi malware, logging dan pemantauan — area penerapan langsung model AI (Bab 4-9)

10.5 Studi Kasus: Kesenjangan Tata Kelola AI pada Insiden Keamanan

Konteks dan Permasalahan

Laporan IBM Cost of a Data Breach 2025 yang telah dibahas pada Bab 2 mengungkap temuan yang sangat relevan bagi bab ini: 97% organisasi yang mengalami insiden keamanan terkait AI ternyata tidak memiliki kontrol akses AI yang memadai (IBM Security & Ponemon Institute, 2025; ThreatLocker Blog, 2026). Temuan lain yang senada mencatat bahwa 63% organisasi tidak memiliki kebijakan tata kelola untuk mengelola AI, termasuk shadow AI yang digunakan karyawan tanpa sepengetahuan tim keamanan (Expert Insights, 2026).

Analisis melalui Lensa NIST CSF 2.0

Ditinjau melalui kerangka NIST CSF 2.0, temuan ini pada dasarnya menunjukkan kegagalan pada fungsi Govern dan Protect: organisasi mengadopsi kapabilitas AI (seringkali dengan tergesa-gesa demi keunggulan kompetitif) tanpa terlebih dahulu membangun kebijakan tata kelola dan kontrol akses yang memadai. Kesenjangan inilah yang mendorong penambahan fungsi Govern sebagai elemen sentral pada CSF 2.0 bukan sebagai formalitas administratif, melainkan sebagai respons langsung terhadap pola kegagalan nyata yang terus berulang di lapangan (SaltyCloud / Isora GRC, 2026; Expert Insights, 2026).

Pembelajaran

Studi kasus ini menggarisbawahi pesan penting bagi buku ini secara keseluruhan: kapabilitas teknis AI yang telah dibahas pada Bagian II dan III sekuat apa pun tidak akan efektif tanpa fondasi tata kelola yang memadai. Integrasi dengan standar seperti NIST CSF 2.0 dan ISO/IEC 27001 bukan sekadar kebutuhan kepatuhan administratif, melainkan prasyarat mendasar agar sistem rekomendasi mitigasi berbasis AI (Bab 9) dapat dipercaya dan diadopsi secara bertanggung jawab oleh organisasi.

Catatan: Prinsip Integrasi Domain Knowledge yang Efektif

- *Rekomendasi mitigasi yang dihasilkan sistem AI sebaiknya secara eksplisit merujuk kontrol standar yang relevan (mis. "tindakan ini memenuhi kontrol A.8.16 ISO 27001 pemantauan aktivitas"), sehingga memudahkan pelaporan kepatuhan.*
- *Fungsi Govern pada NIST CSF 2.0 menuntut keterlacakan penuh atas keputusan AI prinsip ini harus tertanam sejak tahap perancangan sistem, bukan ditambahkan belakangan sebagai lapisan kepatuhan.*

- *Integrasi standar internasional membantu memastikan bahwa rekomendasi mitigasi bersifat praktis dan dapat diterapkan (actionable), bukan sekadar keluaran teknis dari model yang sulit dipetakan ke kebijakan organisasi nyata.*

10.6 Rangkuman

NIST CSF 2.0 menambahkan fungsi Govern sebagai fungsi lintas yang menaungi Identify, Protect, Detect, Respond, dan Recover, mencerminkan pergeseran keamanan siber menjadi risiko tingkat perusahaan. Draf NIST Cyber AI Profile memperluas CSF 2.0 secara khusus untuk risiko dan peluang AI, dengan tiga area fokus: Secure, Defend, dan Thwart. ISO/IEC 27001:2022 menyediakan kontrol teknis rinci pada empat kategori Annex A yang dapat dipetakan langsung terhadap kapabilitas sistem AI yang dibahas di sepanjang buku ini. Studi kasus kesenjangan tata kelola AI menunjukkan bahwa 97% insiden terkait AI terjadi pada organisasi tanpa kontrol akses memadai, menggarisbawahi pentingnya fondasi tata kelola sebelum penerapan teknis.

Daftar Pustaka

SaltyCloud / Isora GRC. (2026). NIST CSF 2.0: What Changed + Free Controls Sheet.

Expert Insights. (2026). NIST Cybersecurity Framework 2.0: What Changed And Why It Matters.

KPMG. (2026). Cybersecurity: NIST Draft Cybersecurity Framework for AI (Cyber AI Profile).

IBM Security & Ponemon Institute. (2025). Cost of a Data Breach Report 2025.

ThreatLocker Blog. (2026). NIST CSF 2.0: How the Framework Is Evolving for Modern Cyber Risk.
International Organization for Standardization. (2022). ISO/IEC 27001:2022 — Information Security Management Systems Requirements.

BAB 11

Studi Kasus : Intelligent Cybersecurity Mitigation Recommendation System Berbasis GNN

11.1 Pendahuluan

Bab ini menyajikan studi kasus implementasi nyata yang menjadi puncak Bagian III buku ini: Intelligent Cybersecurity Mitigation Recommendation System, sebuah sistem rekomendasi mitigasi serangan siber berbasis Graph Neural Networks yang dirancang dan dikembangkan penulis. Studi kasus ini mengintegrasikan hampir seluruh konsep yang telah dibahas pada bab-bab sebelumnya: representasi graf dan GNN (Bab 6), prinsip rekomendasi mitigasi adaptif (Bab 9), serta integrasi dengan standar keamanan internasional (Bab 10).

11.2 Latar Belakang dan Urgensi

Sebagaimana dibahas pada Bab 1 dan Bab 2, serangan siber semakin kompleks dan dinamis, sementara pendekatan tradisional dalam deteksi dan mitigasi kerap gagal menghadapi pola serangan baru yang terus berevolusi. Graph Neural Networks dipilih sebagai pendekatan inti karena kemampuannya memodelkan hubungan antarentitas dalam jaringan

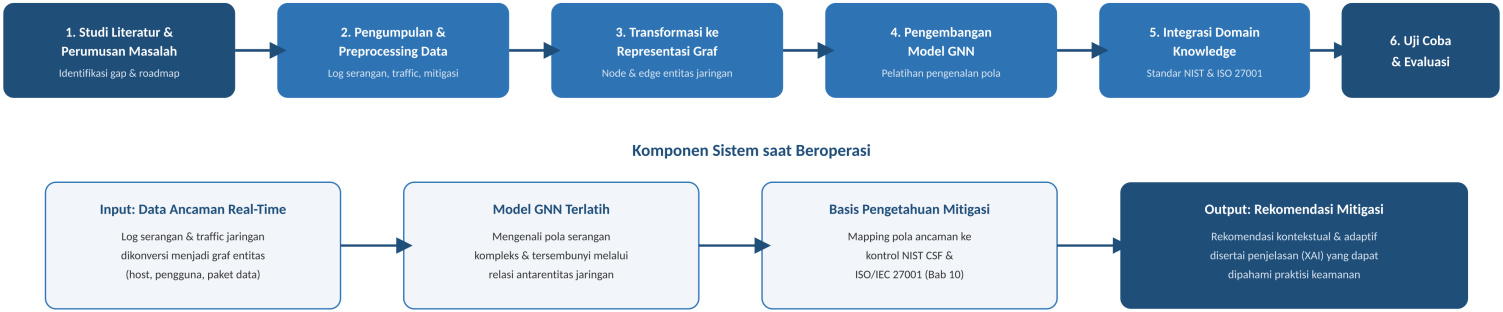
sebagaimana ditegaskan oleh penelitian terbaru yang menunjukkan bahwa GNN dapat meningkatkan efektivitas operasi pertahanan siber dengan mengidentifikasi serangan yang sulit dikenali oleh metode konvensional (Mitra et al., 2024). Tinjauan sistematis lain turut menegaskan bahwa GNN memiliki potensi besar dalam mendeteksi serangan berbahaya secara lebih akurat dibandingkan algoritma tradisional (Alshehri et al., 2025), sementara integrasi GNN dengan explainable AI dan large language model. Dua topik yang telah dibahas pada Bab 7 dinilai mampu memperkuat kemampuan sistem dalam memberikan rekomendasi mitigasi yang transparan dan dapat dipahami praktisi keamanan (AboulEla & Kashef, 2025).

Berdasarkan latar belakang tersebut, studi kasus ini dirumuskan untuk menjawab tiga pertanyaan utama: (1) bagaimana merancang sistem rekomendasi mitigasi serangan siber berbasis GNN yang adaptif; (2) bagaimana mengintegrasikan data ancaman dan strategi mitigasi ke dalam model rekomendasi cerdas; serta (3) sejauh mana sistem yang dikembangkan mampu meningkatkan efektivitas mitigasi dibandingkan metode tradisional.

11.3 Arsitektur Sistem

Gambar 11.1 menyajikan arsitektur menyeluruh sistem, mencakup tahapan pengembangan (bagian atas) dan komponen sistem saat beroperasi (bagian bawah).

Gambar 11.1 Arsitektur Intelligent Cybersecurity Mitigation Recommendation System Berbasis GNN



Sumber: Rancangan penelitian penulis, "Intelligent Cybersecurity Mitigation Recommendation System: Pendekatan Adaptif dengan Graph Neural Networks" (Proposal Penelitian Dasar, 2026).

Gambar 11.1 Arsitektur Intelligent Cybersecurity Mitigation Recommendation System Berbasis GNN

Data ancaman siber berupa log serangan dan traffic jaringan dikumpulkan dan diproses melalui normalisasi, labeling, dan transformasi ke dalam bentuk graf, dengan node merepresentasikan entitas (host, pengguna, paket) dan edge menggambarkan hubungan atau interaksi antarentitas, mengikuti prinsip representasi graf yang telah dibahas pada Bab 6. Model GNN kemudian dilatih untuk mengenali pola serangan yang kompleks dan menghasilkan rekomendasi mitigasi yang kontekstual dan adaptif, sesuai jenis ancaman yang terdeteksi.

11.4 Metode Penelitian dan Peran Tim

Metode penelitian dirancang secara bertahap dan kolaboratif, melibatkan tiga bidang keahlian: Sistem Informasi, Teknologi Informasi, dan Cybersecurity. Tabel 11.1 merinci setiap tahapan penelitian beserta proses, luaran, indikator capaian, dan penanggung jawabnya.

Tabel 11.1 Tahapan Metode Penelitian Studi Kasus

Tahapan	Proses	Luaran	Indikator Capaian	Penanggung Jawab
1. Studi Literatur & Perumusan Masalah	Kajian pustaka, identifikasi gap, rumusan tujuan & roadmap	Kerangka konseptual sistem rekomendasi mitigasi berbasis GNN	Tersusunnya rumusan masalah dan roadmap penelitian	Ketua (Sistem Informasi)
2. Pengumpulan & Preprocessing Data	Pengumpulan log serangan, traffic jaringan, dan strategi mitigasi	Dataset ancaman siber terstruktur dan siap pakai	Volume dan keragaman data yang memadai	Anggota TI & Cybersecurity

3. Transformasi Data ke Graf	Konversi data ke struktur graf (node & edge)	Representasi graf dari data serangan siber	Kualitas struktur graf untuk pelatihan model	Anggota TI
4. Pengembangan Model GNN	Pelatihan model GNN untuk mengenali pola serangan & mitigasi	Model GNN adaptif terlatih	Akurasi dan performa model	Anggota TI
5. Integrasi Domain Knowledge	Memasukkan standar mitigasi (NIST, ISO 27001) ke sistem rekomendasi	Sistem rekomendasi yang sesuai praktik keamanan	Kesesuaian rekomendasi dengan standar industri	Anggota Cybersecurity
6. Uji Coba & Evaluasi Sistem	Simulasi serangan, pengujian sistem, analisis efektivitas mitigasi	Laporan performa sistem & perbandingan dengan metode tradisional	Efektivitas sistem dalam merespons ancaman	Ketua & Anggota Cybersecurity

Integrasi domain knowledge cybersecurity dilakukan dengan memasukkan standar mitigasi internasional seperti NIST dan ISO 27001 sebagaimana dibahas mendalam pada Bab 10 agar rekomendasi yang dihasilkan sesuai dengan praktik keamanan global (NIST, 2020; ISO/IEC, 2018). Tahap akhir penelitian adalah uji coba sistem melalui simulasi serangan dan evaluasi efektivitas mitigasi, dengan indikator berupa akurasi sistem, kesesuaian rekomendasi, serta perbandingan performa dengan metode tradisional.

11.5 Roadmap Penelitian: Kesinambungan Kompetensi Peneliti

Studi kasus ini bukan merupakan lompatan riset yang berdiri sendiri, melainkan kelanjutan logis dari perjalanan riset ketua tim peneliti pada bidang sistem rekomendasi dan analisis data selama beberapa tahun terakhir. Tabel 11.2 menggambarkan kesinambungan tersebut.

Tabel 11.2 Roadmap Penelitian dan Kesinambungan Kompetensi

Tahun	Fokus Penelitian Ketua Tim	Kontribusi terhadap Studi Kasus 2026
2023	Clustering pelaksanaan vaksinasi di Jawa Tengah menggunakan metode K-Means	Fondasi kompetensi unsupervised learning untuk analisis pola data
2024	Content-based filtering pada sistem rekomendasi buku informatika	Fondasi kompetensi desain algoritma sistem rekomendasi
2025	Sistem rekomendasi makanan kucing menggunakan content-based filtering	Penguatan keterampilan desain rekomendasi pada domain berbeda
2026 (studi kasus ini)	Intelligent Cybersecurity Mitigation Recommendation System berbasis GNN	Integrasi kompetensi rekomendasi + representasi graf untuk domain keamanan siber
2027 (rencana lanjutan)	Optimasi sistem dengan Explainable AI	Penambahan transparansi agar rekomendasi dapat dipahami praktisi keamanan (Bab 7)

Keterkaitan antara tahapan penelitian 2023–2025 dengan studi kasus 2026 menunjukkan evolusi metodologis dan penguatan kompetensi peneliti

dalam sistem rekomendasi dan analisis data, yang menjadi fondasi kuat bagi pengembangan sistem mitigasi siber berbasis GNN. Rencana lanjutan pada 2027 optimasi sistem dengan explainable AI akan menerapkan langsung prinsip-prinsip yang telah dibahas pada Bab 7 buku ini.

11.6 Luaran yang Ditargetkan

Tabel 11.3 merangkum luaran yang ditargetkan dari studi kasus ini, mencakup dimensi akademik maupun praktis.

Tabel 11.3 Luaran yang Ditargetkan dari Studi Kasus

Jenis Luaran	Deskripsi
Prototipe sistem	Sistem rekomendasi mitigasi serangan siber berbasis Graph Neural Networks yang adaptif dan kontekstual
Dataset terstruktur	Dataset ancaman siber yang telah direpresentasikan dalam bentuk graf, dapat dimanfaatkan untuk riset lanjutan
Model terlatih	Model GNN yang dapat digunakan kembali untuk pengembangan sistem keamanan lainnya
Publikasi ilmiah	Artikel pada jurnal terakreditasi nasional (SINTA 1-4) dan seminar nasional/internasional
Modul pelatihan	Materi peningkatan literasi keamanan siber bagi mahasiswa, komunitas, dan praktisi
Hak cipta	Pendaftaran hak cipta atas prototipe sistem yang dikembangkan

Catatan: Kebaruan dan Posisi Studi Kasus dalam Lanskap Riset

- *Sebagian besar penelitian terdahulu di bidang keamanan siber berfokus pada deteksi intrusi menggunakan machine learning konvensional atau rule-based system yang cenderung statis (Bab 2 & Bab 4); studi kasus ini melangkah lebih jauh dengan menghasilkan rekomendasi mitigasi adaptif, bukan sekadar deteksi.*
- *Kebaruan utama terletak pada kemampuan GNN memodelkan hubungan antarentitas jaringan secara dinamis (Bab 6), dikombinasikan dengan domain knowledge cybersecurity dan standar keamanan internasional (Bab 10), sehingga rekomendasi yang dihasilkan bersifat praktis dan dapat diterapkan.*
- *Rencana integrasi Explainable AI pada tahap lanjutan (2027) akan memastikan rekomendasi yang dihasilkan dapat dipahami dan dipercaya oleh praktisi keamanan, sejalan dengan prinsip yang dibahas pada Bab 7.*
- *Studi kasus ini mendemonstrasikan secara konkret bagaimana seluruh kerangka konseptual yang dibangun sepanjang buku ini mulai dari lanskap ancaman (Bab 1) hingga standar keamanan (Bab 10) dapat disatukan menjadi satu sistem yang koheren dan siap diuji di lapangan.*

11.7 Purwarupa Sistem: Implementasi pada Platform Web

Sebagai bukti kelayakan (proof of concept) dari rancangan pada Subbab 11.3–11.4, sistem yang dibahas pada bab ini telah diimplementasikan menjadi purwarupa (prototype) berbasis web yang dapat diakses publik pada alamat deteksi-cyber.kuliahtsu.my.id. Berbeda dengan sebagian besar studi yang dirujuk pada bab-bab sebelumnya buku ini yang umumnya divalidasi pada dataset benchmark standar seperti NSL-KDD atau

CICIDS2017 (Bab 4) purwarupa ini divalidasi menggunakan data serangan nyata yang dikumpulkan dari infrastruktur honeypot milik penulis sendiri, memberikan gambaran yang lebih otentik mengenai bagaimana sistem akan berperilaku pada kondisi operasional sesungguhnya.

11.7.1 Sumber Data dan Arsitektur Purwarupa

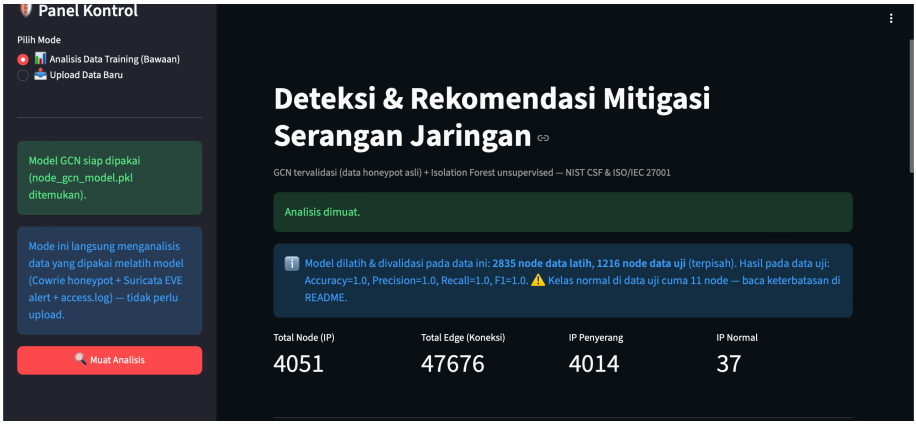
Purwarupa dibangun menggunakan kerangka kerja Streamlit dan menyediakan dua mode operasi: mode "Analisis Data Training (Bawaan)", yang langsung menganalisis data yang telah digunakan untuk melatih model tanpa memerlukan unggahan data tambahan, dan mode "Upload Data Baru", yang memungkinkan pengguna mengunggah data trafik baru untuk dianalisis oleh model yang telah dilatih. Data pelatihan bersumber dari kombinasi tiga jenis log nyata: Cowrie honeypot (mensimulasikan layanan SSH/Telnet rentan untuk memancing penyerang), Suricata EVE alert (log deteksi intrusi berbasis signature sebagai pembanding, Bab 2), dan access log server web mengikuti prinsip pengumpulan data multi-sumber yang telah dibahas pada Bab 6 (Subbab 6.3.1).

Model inti yang digunakan adalah Graph Convolutional Network (node_gcn_model.pkl), diperkuat dengan Isolation Forest sebagai lapisan deteksi anomali tak terawasi tambahan sebuah kombinasi supervised-unsupervised yang selaras dengan prinsip deteksi berlapis yang telah dibahas pada Bab 4 (Pseudocode 4.4, Random Forest-Autoencoder). Tabel 11.4 merangkum statistik graf jaringan yang dihasilkan dari data honeypot penulis.

Tabel 11.4 Statistik Graf Jaringan pada Purwarupa (Data Honeypot Nyata)

Metrik	Nilai
Total node (alamat IP unik)	4.051

Total edge (koneksi tercatat)	47.676
IP teridentifikasi sebagai penyerang	4.014 (99,1% dari total node)
IP teridentifikasi sebagai normal	37 (0,9% dari total node)
Jumlah node data latih / data uji	2.835 node latih / 1.216 node uji (terpisah/hold-out)
Akurasi, Presisi, Recall, F1, AUC-ROC pada data uji	1,0 / 1,0 / 1,0 / 1,0 / 1,0



Gambar 11.2 Tampilan Panel Ringkasan Purwarupa deteksi-cyber.kuliahtsu.my.id

11.7.2 Catatan Kejujuran Ilmiah: Keterbatasan Metrik pada Data Uji

Sejalan dengan prinsip kehati-hatian ilmiah yang perlu ditekankan pada seluruh buku ini, penting dicatat bahwa metrik evaluasi sempurna (1,0 pada seluruh indikator) yang ditampilkan pada dashboard purwarupa perlu diinterpretasikan secara hati-hati. Purwarupa secara eksplisit menampilkan peringatan bahwa kelas "normal" pada data uji hanya berjumlah 11 node dari total 1.216 node uji sebuah kondisi ketidakseimbangan kelas (class imbalance) yang ekstrem, sebagaimana telah diperingatkan pada Bab 4

(Subbab 4.2.7). Pada kondisi seperti ini, metrik akurasi maupun F1-score yang tampak sempurna dapat menyesatkan, karena model berpotensi mencapai skor tinggi hanya dengan mengenali pola mayoritas (trafik penyerang) dengan sangat baik, sementara kemampuannya mengenali kelas minoritas (trafik normal) yang jumlahnya sangat sedikit belum tentu teruji secara memadai.

Transparansi purwarupa dalam menampilkan peringatan ini secara langsung pada antarmuka penggunanya merupakan praktik yang patut dicontoh, dan sejalan dengan prinsip explainable AI yang dibahas pada Bab 7: pengguna sistem baik peneliti maupun praktisi keamanan yang mengevaluasi kelayakan adopsi — berhak mengetahui keterbatasan model, bukan hanya angka performa yang tampak mengesankan. Karakteristik data ini juga mencerminkan realitas operasional honeypot: karena honeypot dirancang khusus untuk memancing serangan, proporsi trafik yang tercatat memang secara alami didominasi oleh aktivitas mencurigakan, sehingga pengembangan lanjutan sistem ini perlu mempertimbangkan augmentasi data trafik normal dari sumber tambahan agar evaluasi model menjadi lebih representatif.

Rekomendasi Metrik yang Lebih Tepat untuk Data Tidak Seimbang

Sebagaimana disinggung pada Subbab 4.2.7, akurasi dan bahkan F1-score standar (yang dihitung sebagai rata-rata makro atau mikro di seluruh kelas) dapat menyembunyikan performa buruk pada kelas minoritas ketika satu kelas mendominasi data secara ekstrem seperti pada kasus purwarupa ini (4.014 IP penyerang berbanding 37 IP normal). Untuk melaporkan performa model secara lebih objektif pada kondisi seperti ini, dua pendekatan berikut lebih disarankan dibandingkan mengandalkan akurasi keseluruhan semata:

1. **F1-score per kelas (per-class F1)**. Alih-alih melaporkan satu angka F1 gabungan, hitung dan laporkan F1-score secara terpisah untuk kelas "normal" dan kelas "penyerang". Pada kasus purwarupa ini, F1-score untuk kelas penyerang (dengan 1.205 sampel uji) hampir pasti akan sangat tinggi mengingat jumlah sampelnya yang besar, namun F1-score untuk kelas normal (hanya 11 sampel uji) jauh lebih rentan berfluktuasi dan berpotensi menunjukkan gambaran performa yang sangat berbeda — bisa jadi model hanya berhasil mengenali sebagian kecil dari 11 sampel normal tersebut meski F1 gabungan tetap tampak sempurna.
2. **Precision-Recall AUC (PR-AUC).** Berbeda dengan ROC-AUC yang dapat tetap tampak baik meski performa pada kelas minoritas buruk (karena ROC-AUC turut memperhitungkan true negative rate yang mendominasi ketika kelas mayoritas sangat besar), PR-AUC berfokus secara khusus pada trade-off antara presisi dan recall untuk kelas positif (dalam konteks ini, dapat didefinisikan sebagai kelas "normal" yang justru menjadi kelas minoritas dan lebih sulit dideteksi). PR-AUC yang rendah meski ROC-AUC tinggi merupakan indikator yang jauh lebih jujur bahwa model kesulitan mengenali kelas minoritas.

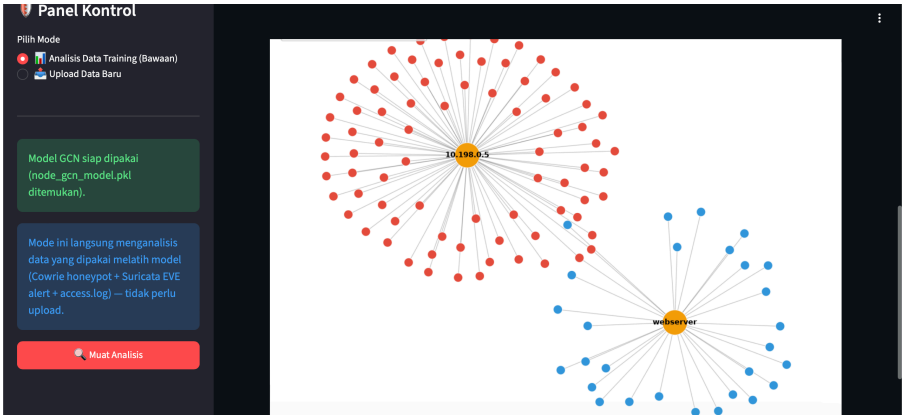
Sebagai ilustrasi mengapa kedua metrik ini penting: bayangkan sebuah model "malas" yang secara sederhana selalu memprediksi setiap node sebagai "penyerang" tanpa benar-benar mempelajari pola apa pun. Pada data uji purwarupa ini, model malas tersebut akan tetap mencapai akurasi sebesar 1.205/1.216 ≈ 99,1% angka yang tampak sangat mengesankan namun sepenuhnya tidak berguna secara praktis, karena model tersebut gagal total mendeteksi satu pun kasus normal. F1-score per kelas dan PR-AUC akan segera mengungkap kegagalan ini F1-score untuk kelas normal akan bernilai 0, sementara PR-AUC untuk kelas normal juga akan mendekati batas bawah sementara akurasi keseluruhan maupun F1-score gabungan

berpotensi tetap tampak tinggi. Pengembangan lanjutan purwarupa ini sebaiknya menyertakan pelaporan kedua metrik tersebut secara eksplisit pada dashboard, alih-alih hanya menampilkan metrik akurasi, presisi, recall, dan F1 gabungan yang saat ini ditampilkan (lihat Gambar 11.2).

Pembelajaran ini memiliki nilai pedagogis yang penting bagi pembaca buku ini, khususnya mahasiswa dan peneliti pemula: angka performa yang tampak sempurna harus selalu memicu kewaspadaan, bukan kepuasan pertanyaan pertama yang perlu diajukan adalah "seberapa seimbang distribusi kelas pada data uji ini?", dan jika jawabannya menunjukkan ketidakseimbangan signifikan, metrik agregat seperti akurasi perlu dilengkapi (bukan digantikan) dengan metrik per kelas seperti F1 per kelas dan PR-AUC agar gambaran performa model dilaporkan secara utuh dan tidak menyesatkan.

11.7.3 Visualisasi Graf dan Identifikasi Pola Serangan

Purwarupa menyediakan visualisasi graf jaringan interaktif yang secara langsung menerapkan prinsip representasi graf pada Bab 6 (Gambar 6.1). Gambar 11.3 menampilkan contoh visualisasi yang dihasilkan sistem, memperlihatkan struktur star-topology di sekitar dua entitas: sebuah host gateway (10.198.0.5) dan sebuah node webserver, masing-masing menunjukkan puluhan koneksi masuk dari alamat IP yang berbeda-beda.



Gambar 11.3 Visualisasi Graf Jaringan pada Purwarupa, Memperlihatkan Pola Koneksi di Sekitar Host Gateway dan Webserver

Pola star-topology semacam ini satu node pusat dengan banyak koneksi radial dari node lain merupakan salah satu indikator visual yang dapat membantu analis mengidentifikasi target yang menjadi sasaran utama percobaan intrusi (dalam hal ini, gateway dan webserver), sekaligus mendemonstrasikan secara konkret keunggulan representasi graf yang telah dibahas pada Bab 6: pola relasional semacam ini akan sulit dikenali apabila data hanya dianalisis sebagai baris-baris tabel independen, sebagaimana dilakukan pendekatan machine learning konvensional pada Bab 4.

11.7.4 Identifikasi Penyerang dan Rekomendasi Mitigasi Kontekstual

Menerapkan langsung prinsip sistem rekomendasi mitigasi adaptif yang dibahas pada Bab 9, purwarupa menghasilkan daftar peringkat IP penyerang paling aktif beserta rekomendasi mitigasi yang dipetakan terhadap kontrol NIST CSF dan ISO/IEC 27001 (Bab 10). Tabel 11.5 menyajikan contoh nyata keluaran sistem untuk lima IP penyerang teratas berdasarkan jumlah percobaan koneksi.

Tabel 11.5 Contoh Keluaran Rekomendasi Mitigasi dari Purwarupa (Data Nyata)

IP Penyerang	Percobaan Koneksi	Tingkat Keparahan	Rekomendasi Mitigasi (Ringkas)	Referensi NIST CSF / ISO 27001
143.110.189.170	1.379	Tinggi	Terapkan fail2ban/rate limiting SSH; nonaktifkan autentikasi password (SSH key-only); pindahkan port SSH dari default	PR.AC-7, DE.CM-3 / A.9.4.2, A.9.4.3
143.110.189.215	822	Tinggi	Terapkan fail2ban/rate limiting SSH; nonaktifkan autentikasi password (SSH key-only); pindahkan port SSH dari default	PR.AC-7, DE.CM-3 / A.9.4.2, A.9.4.3
170.64.175.213	471	Tinggi	Terapkan fail2ban/rate limiting SSH; nonaktifkan autentikasi password (SSH key-only); pindahkan port SSH dari default	PR.AC-7, DE.CM-3 / A.9.4.2, A.9.4.3
79.13.248.116	369	Tinggi	Terapkan fail2ban/rate limiting SSH; nonaktifkan autentikasi password (SSH key-only); pindahkan port SSH dari default	PR.AC-7, DE.CM-3 / A.9.4.2, A.9.4.3

77.90.185.122	115	Tinggi	Terapkan fail2ban/rate limiting SSH; nonaktifkan autentikasi password (SSH key-only); pindahkan port SSH dari default	PR.AC-7, DE.CM-3 / A.9.4.2, A.9.4.3

Gambar 11.4 Tampilan Tab Rekomendasi Mitigasi pada Purwarupa, Menunjukkan Pemetaan terhadap NIST CSF dan ISO/IEC 27001

Pola pada Tabel 11.5 secara konsisten mengarah pada rekomendasi penguatan keamanan layanan SSH konsisten dengan sifat honeypot Cowrie yang secara khusus dirancang mensimulasikan layanan tersebut. Yang secara khusus relevan bagi pembahasan Bab 10, setiap rekomendasi secara otomatis dipetakan terhadap kontrol PR.AC-7 (Access Control) dan DE.CM-3 (Personnel Activity Monitoring) pada NIST CSF, serta kontrol A.9.4.2 (Secure Log-on Procedures) dan A.9.4.3 (Password Management) pada ISO/IEC 27001 mendemonstrasikan secara konkret bagaimana keluaran mesin rekomendasi berbasis GNN dapat langsung ditelusuri ke kerangka

kepatuhan yang diakui secara internasional, alih-alih menghasilkan rekomendasi teknis yang berdiri sendiri tanpa keterkaitan tata kelola.

Menariknya, sistem juga mencatat variasi pada kolom "Distinct Port Dicoba" , misalnya IP 77.90.185.122 mencoba 48 port berbeda meski hanya melakukan 115 percobaan koneksi, mengindikasikan pola port scanning, berbeda dengan IP 143.110.189.170 yang melakukan 1.379 percobaan koneksi namun hanya menyasar 1 port, mengindikasikan pola brute-force terfokus. Perbedaan pola inilah yang menjadi dasar bagi pengembangan lanjutan sistem agar dapat menghasilkan rekomendasi yang lebih terdiferensiasi berdasarkan taktik spesifik yang terdeteksi, bukan sekadar rekomendasi generik berdasarkan volume koneksi semata arah pengembangan yang sejalan dengan rencana integrasi Explainable AI pada Subbab 11.5.

Catatan: Nilai Purwarupa sebagai Jembatan Riset-ke-Praktik

- *Purwarupa yang dibangun dari data honeypot nyata (bukan dataset benchmark akademik) memberikan validasi tambahan bahwa arsitektur pada Gambar 11.1 layak diimplementasikan pada kondisi operasional sesungguhnya, melengkapi validasi akademik yang direncanakan pada tahapan penelitian (Tabel 11.1).*
- *Keterbukaan purwarupa dalam menampilkan keterbatasan (class imbalance pada data uji) mencontohkan praktik pelaporan ilmiah yang jujur, sejalan dengan semangat kehati-hatian epistemik yang ditekankan di sepanjang buku ini.*
- *Pemetaan otomatis rekomendasi mitigasi terhadap NIST CSF dan ISO/IEC 27001 pada purwarupa membuktikan bahwa integrasi domain knowledge yang dibahas secara konseptual pada Bab 10*

dapat diimplementasikan secara nyata dalam sistem operasional, bukan sekadar kerangka teoretis.

- *Visualisasi graf interaktif pada purwarupa menunjukkan bagaimana representasi graf (Bab 6) dapat disajikan dengan cara yang intuitif bagi analis keamanan tanpa latar belakang matematis mendalam, mendukung adopsi praktis di lapangan.*

11.8 Rangkuman

Studi kasus ini mengintegrasikan representasi graf, model GNN, dan standar keamanan internasional menjadi satu sistem rekomendasi mitigasi yang adaptif dan kontekstual. Metode penelitian dirancang secara bertahap dan kolaboratif lintas tiga bidang keahlian, dengan indikator capaian yang terukur pada setiap tahapan. Studi kasus merupakan kelanjutan logis dari roadmap riset ketua tim peneliti pada bidang sistem rekomendasi sejak 2023, bukan lompatan riset yang berdiri sendiri. Luaran yang ditargetkan mencakup dimensi akademik (publikasi, model, dataset) maupun praktis (modul pelatihan, hak cipta prototipe). Purwarupa berbasis web (deteksi-cyber.kuliahtsu.my.id) yang divalidasi pada data honeypot nyata membuktikan kelayakan operasional arsitektur sistem, sekaligus mendemonstrasikan integrasi nyata antara deteksi berbasis GNN, rekomendasi mitigasi kontekstual, dan pemetaan terhadap NIST CSF/ISO 27001. Rencana pengembangan lanjutan pada 2027 integrasi Explainable AI akan menerapkan langsung prinsip transparansi yang dibahas pada Bab 7 buku ini.

Daftar Pustaka

Mitra, S., Chakraborty, T., Neupane, S., Piplai, A., & Mittal, S. (2024). Use of Graph Neural Networks in Aiding Defensive Cyber Operations. arXiv:2401.05680.

Alshehri, S. M., Sharaf, S. A., & Molla, R. A. (2025). Systematic Review of Graph Neural Network for Malicious Attack Detection. Information, 16(6), 470.

AboulEla, S. G., & Kashef, R. F. (2025). Leveraging Large Language Models, Graph Neural Networks, and Explainable AI for Next-Generation Network Intrusion Detection Systems. Journal of Intelligent Information Systems, 63, 1807–1835.

National Institute of Standards and Technology (NIST). (2020). Cybersecurity Framework. U.S. Department of Commerce.

ISO/IEC. (2018). ISO/IEC 27001: Information Security Management Systems – Requirements.

Ghosh, U., Paul, A., Sarkar, D., Dey, M., & Paul, P. (2025). Graph-Based Approaches in Cybersecurity: A Comprehensive Survey. Springer.

Penulis. (2026). Purwarupa Intelligent Cybersecurity Mitigation Recommendation System. Diakses dari https://deteksi-cyber.kuliahtsu.my.id.

BAB 12

Tantangan Etis dan Privasi dalam AI untuk Keamanan Siber

12.1 Pendahuluan

Bagian III buku ini telah menunjukkan bagaimana AI dapat secara signifikan memperkuat kemampuan deteksi dan mitigasi ancaman siber. Namun, sebagaimana disinggung sejak Bab 2, kekuatan ini datang bersama tanggung jawab: sistem yang memantau perilaku pengguna, menganalisis data sensitif, dan bahkan mengambil tindakan otonom menimbulkan pertanyaan etis dan privasi yang tidak dapat diabaikan. Bab ini membuka Bagian IV dengan membahas isu-isu tersebut secara khusus.

12.2 Peta Isu Etis dalam AI untuk Keamanan Siber

Tabel 12.1 merangkum lima isu etis utama yang relevan dengan penerapan AI dalam keamanan siber, sebagaimana telah disinggung pada berbagai bab sebelumnya di buku ini.

Tabel 12.1 Peta Isu Etis dalam Penerapan AI untuk Keamanan Siber

Isu Etis	Deskripsi Risiko	Contoh Dampak
Bias Model	Model dilatih pada data historis yang mungkin mengandung bias tersembunyi	Profil risiko yang secara tidak adil menandai kelompok pengguna tertentu sebagai lebih mencurigakan

Privasi Data	Model memerlukan data perilaku pengguna yang sensitif untuk deteksi anomali (Bab 4, 8)	Penyalahgunaan data pemantauan karyawan di luar tujuan keamanan yang sah
Akuntabilitas	Sulit menentukan pihak yang bertanggung jawab atas keputusan otonom AI	Tindakan mitigasi otomatis yang keliru merugikan operasional tanpa kejelasan tanggung jawab
Transparansi (Black Box)	Model kompleks sulit dijelaskan kepada pemangku kepentingan non-teknis	Keputusan yang tidak dapat dipertanggungjawabkan secara memadai kepada regulator atau auditor
Ketergantungan Berlebihan (Automation Bias)	Analis terlalu mempercayai rekomendasi AI tanpa verifikasi kritis	Kegagalan mendeteksi kesalahan model karena kepercayaan berlebihan pada otomasi

12.3 Privasi Data: Ketegangan antara Deteksi dan Perlindungan

Sistem deteksi ancaman orang dalam yang dibahas pada Bab 8 memerlukan pemantauan perilaku karyawan pola akses berkas, aktivitas di luar jam kerja, volume transfer data untuk membangun profil "perilaku normal" yang efektif. Data semacam ini secara inheren bersifat sensitif, dan pengumpulannya menimbulkan ketegangan yang nyata antara kebutuhan keamanan organisasi dan hak privasi individu karyawan. Ketegangan serupa juga muncul pada deteksi anomali transaksi perbankan yang dibahas pada Bab 4, yang bergantung pada analisis pola perilaku finansial nasabah secara mendalam.

Prinsip data minimization mengumpulkan hanya data yang benar-benar diperlukan untuk tujuan keamanan spesifik — dan pembatasan retensi data menjadi praktik penting untuk menyeimbangkan kedua kepentingan tersebut. Pendekatan federated learning yang akan dibahas pada Bab 13 juga menawarkan salah satu solusi teknis terhadap ketegangan ini, dengan memungkinkan pembelajaran kolaboratif tanpa perlu memusatkan data mentah yang sensitif.

12.4 Bias Model dan Keadilan Algoritmik

Model machine learning dan deep learning yang dibahas pada Bagian II buku ini dilatih menggunakan data historis. Jika data tersebut mengandung pola bias misalnya karena praktik pengumpulan data yang tidak representatif, atau karena mencerminkan bias struktural yang sudah ada sebelumnya dalam organisasi model berisiko mereplikasi dan bahkan memperkuat bias tersebut dalam skala yang lebih besar dan dengan kecepatan yang lebih tinggi. Dalam konteks keamanan siber, bias semacam ini dapat bermanifestasi sebagai profil risiko yang secara sistematis menandai kelompok pengguna tertentu sebagai lebih mencurigakan tanpa dasar yang valid, berpotensi menimbulkan konsekuensi yang tidak adil bagi individu yang terdampak.

12.5 Akuntabilitas atas Tindakan Otonom

Bab 8 dan Bab 9 telah membahas bagaimana sistem AI SOC modern dapat mengeksekusi tindakan mitigasi secara otomatis penuh untuk skenario berisiko rendah dan berkeyakinan tinggi. Kapabilitas ini menghadirkan pertanyaan akuntabilitas yang mendasar: ketika tindakan otomatis AI keliru dan menyebabkan gangguan operasional atau kerugian, pihak mana yang bertanggung jawab pengembang model, organisasi yang mengoperasikan

sistem, atau vendor platform AI SOC? Praktik human-in-the-loop yang telah ditekankan pada Bab 8 sebagian merupakan respons terhadap persoalan akuntabilitas ini, dengan mempertahankan keterlibatan manusia pada keputusan berdampak signifikan.

12.6 Keamanan Sistem AI Itu Sendiri: Adversarial Machine Learning

Bagian ini membahas dimensi yang secara khusus relevan bagi buku dengan fokus "keamanan siber": bahwa model AI yang telah dibahas di sepanjang Bagian II dan III buku ini bukan hanya alat pertahanan, melainkan juga permukaan serangan (attack surface) baru yang perlu diamankan. Penyerang modern tidak lagi terbatas pada mengeksploitasi kerentanan jaringan atau sistem operasi sebagaimana dibahas pada Bab 1–2, tetapi juga dapat secara langsung menargetkan model machine learning, deep learning, dan GNN yang menjadi inti sistem deteksi dan rekomendasi mitigasi yang dibahas buku ini. Bidang kajian yang mempelajari kerentanan dan pertahanan model AI terhadap serangan semacam ini dikenal sebagai adversarial machine learning.

12.6.1 Adversarial Perturbation: Mengelabui Model Saat Inferensi

Adversarial perturbation adalah teknik memodifikasi data masukan secara halus dan seringkali tidak kasatmata, dengan tujuan mengecoh model agar menghasilkan prediksi yang keliru, sementara data yang dimodifikasi tetap tampak normal bagi pengamat manusia atau bagi sistem pemantauan lain. Pada konteks deteksi intrusi berbasis machine learning (Bab 4), penyerang dapat secara sengaja menyesuaikan pola trafik jaringan mereka, misalnya mengatur ulang waktu antarpaket, memecah payload berbahaya menjadi potongan yang lebih kecil, atau menambahkan trafik "pengisi" yang tidak

berbahaya. Sedemikian rupa sehingga vektor fitur yang dihasilkan berada tepat di luar batas keputusan (decision boundary) model, sehingga trafik berbahaya tersebut diklasifikasikan sebagai normal.

Secara konseptual, serangan semacam ini memanfaatkan kenyataan bahwa model machine learning mempelajari batas keputusan berdasarkan data pelatihan yang terbatas, bukan pemahaman semantik yang sesungguhnya tentang apa yang membuat suatu trafik berbahaya. Perturbasi sekecil apa pun pada fitur masukan, asalkan cukup untuk melewati batas keputusan tersebut dapat mengubah hasil klasifikasi secara drastis, meskipun perubahan pada data asli tergolong minor. Kerentanan ini telah didokumentasikan secara luas pada domain computer vision, di mana perubahan piksel yang tidak kasatmata dapat mengecoh model klasifikasi citra (Goodfellow et al., 2015), dan kini menjadi perhatian yang semakin besar pada domain keamanan siber, mengingat konsekuensi kegagalan deteksi jauh lebih serius dibandingkan kesalahan klasifikasi citra semata.

12.6.2 Data Poisoning: Meracuni Model Sejak Tahap Pelatihan

Berbeda dengan adversarial perturbation yang menyerang model pada tahap inferensi (saat model sudah digunakan), data poisoning menargetkan tahap pelatihan model itu sendiri. Penyerang yang memiliki akses, baik langsung maupun tidak langsung terhadap data yang digunakan untuk melatih atau melatih ulang (retraining) model, dapat menyisipkan sampel data yang telah dimanipulasi secara sengaja, dengan tujuan menanamkan bias atau "pintu belakang" (backdoor) tersembunyi pada model yang dihasilkan.

Kerentanan ini secara khusus relevan bagi sistem yang menerapkan pembelajaran berkelanjutan (continuous learning) sebuah praktik yang telah direkomendasikan pada Bab 4 dan Bab 9 sebagai cara menjaga model

tetap adaptif terhadap pola ancaman baru. Jika mekanisme umpan balik yang digunakan untuk melatih ulang model tidak divalidasi secara memadai, penyerang berpotensi secara bertahap "meracuni" model dengan menyisipkan data yang tampak sah namun sesungguhnya dirancang untuk melemahkan kemampuan deteksi terhadap pola serangan tertentu di masa depan. Skenario ini juga relevan bagi pendekatan federated learning yang dibahas pada Bab 13: karena setiap node melatih model secara lokal, node yang disusupi (compromised) berpotensi mengirimkan pembaruan model yang telah diracuni ke dalam proses agregasi global, berpotensi memengaruhi model gabungan yang digunakan seluruh organisasi peserta.

12.6.3 Strategi Pertahanan terhadap Adversarial Machine Learning

Sejumlah strategi pertahanan telah dikembangkan untuk memitigasi risiko adversarial machine learning, meskipun bidang ini masih menjadi area riset yang sangat aktif dan belum menawarkan solusi yang sepenuhnya tuntas (Biggio & Roli, 2018):

1. Adversarial training. Melatihmodel tidak hanya pada data normal, tetapi juga menyertakan sampel adversarial yang dihasilkan secara sengaja selama proses pelatihan, sehingga model belajar mengenali dan tetap robust terhadap pola perturbasi yang umum digunakan penyerang.
2. Validasi dan sanitasi data pelatihan. Menerapkan pemeriksaan anomali terhadap data yang akan digunakan untuk melatih ulang model (termasuk data yang berasal dari mekanisme umpan balik operasional), guna mendeteksi dan menyaring sampel yang berpotensi merupakan upaya data poisoning sebelum data tersebut memengaruhi model.

3. Ensemble dan diversifikasi model. Menggunakan kombinasi beberapa model dengan arsitektur atau data pelatihan yang berbeda (mengingat prinsip ensemble yang telah dibahas pada Bab 4), sehingga kerentanan yang menargetkan satu arsitektur spesifik tidak serta-merta melumpuhkan keseluruhan sistem deteksi.
4. Pemantauan drift model. Memantau performa model secara berkelanjutan pada lingkungan produksi untuk mendeteksi penurunan akurasi yang tidak wajar atau pola prediksi yang mencurigakan, yang dapat menjadi indikasi awal keberhasilan serangan data poisoning atau adversarial.

Relevansi pembahasan ini bagi buku secara keseluruhan sangat mendasar: seluruh kapabilitas deteksi dan rekomendasi mitigasi berbasis AI yang dibahas pada Bagian II dan III. Mulai dari machine learning klasik (Bab 4), deep learning (Bab 5), hingga graph neural network (Bab 6) pada akhirnya hanya akan sekuat ketahanan model itu sendiri terhadap upaya manipulasi yang disengaja. Keamanan siber berbasis AI, dengan demikian, bukan hanya soal menggunakan AI untuk mengamankan sistem, melainkan juga soal mengamankan AI itu sendiri sebagai bagian integral dari permukaan pertahanan organisasi.

12.7 Studi Kasus: Kesenjangan Tata Kelola AI dan Risiko Shadow AI

Konteks dan Permasalahan

Melanjutkan temuan yang telah disinggung pada Bab 2 dan Bab 10, laporan IBM Cost of a Data Breach 2025 mengungkap bahwa 97% organisasi yang mengalami insiden keamanan terkait AI tidak memiliki kontrol akses AI yang memadai (IBM Security & Ponemon Institute, 2025). Temuan yang lebih spesifik mengenai fenomena shadow AI penggunaan alat AI oleh karyawan

tanpa persetujuan atau pengawasan tim keamanan menunjukkan bahwa insiden yang melibatkan shadow AI memakan waktu deteksi dan penahanan yang lebih lama (247 hari) dibandingkan rata-rata keseluruhan (241 hari), serta menimbulkan biaya tambahan yang signifikan bagi organisasi (IBM Security & Ponemon Institute, 2025; DataFence, 2025).

Analisis dan Pembelajaran

Fenomena shadow AI merupakan ilustrasi konkret dari beberapa isu etis yang dibahas pada bab ini sekaligus: kurangnya akuntabilitas yang jelas atas penggunaan AI oleh individu karyawan, risiko privasi data yang mengalir ke layanan AI eksternal tanpa pengawasan, serta kesenjangan tata kelola yang telah dibahas pada Bab 10 melalui lensa fungsi Govern pada NIST CSF 2.0. Studi ini menunjukkan bahwa tantangan etis dalam AI untuk keamanan siber bukan sekadar wacana filosofis, melainkan risiko operasional nyata dengan konsekuensi finansial dan operasional yang terukur.

Catatan: Prinsip Penerapan AI yang Bertanggung Jawab dalam Keamanan Siber

- *Terapkan prinsip data minimization dan pembatasan retensi data pada seluruh sistem pemantauan berbasis AI, khususnya untuk data perilaku individu yang sensitif.*
- *Lakukan audit bias secara berkala terhadap model yang digunakan untuk profil risiko, dengan melibatkan tinjauan lintas-disiplin, bukan hanya evaluasi teknis semata.*
- *Pertahankan human-in-the-loop untuk keputusan berdampak signifikan, sejalan dengan prinsip yang telah dibahas pada Bab 8, sebagai bentuk mitigasi risiko akuntabilitas.*

- *Bangun kebijakan tata kelola AI yang eksplisit dan sosialisasikan secara luas kepada seluruh karyawan untuk mencegah risiko shadow AI, selaras dengan fungsi Govern pada NIST CSF 2.0 (Bab 10).*
- *Perlakukan model AI itu sendiri sebagai aset yang perlu diamankan, terapkan adversarial training, validasi data pelatihan, dan pemantauan drift model sebagai bagian rutin siklus hidup sistem AI (Subbab 12.6.3).*

12.8 Rangkuman

Penerapan AI dalam keamanan siber menimbulkan enam isu etis dan keamanan utama: bias model, privasi data, akuntabilitas, transparansi, ketergantungan berlebihan pada otomasi, serta kerentanan model AI itu sendiri terhadap manipulasi. Ketegangan antara kebutuhan deteksi ancaman dan perlindungan privasi individu menuntut penerapan prinsip data minimization dan solusi teknis seperti federated learning. Bias model berisiko mereplikasi dan memperbesar bias struktural yang ada dalam data historis, berpotensi menimbulkan profil risiko yang tidak adil. Model AI itu sendiri merupakan permukaan serangan baru: adversarial perturbation dapat mengecoh model pada tahap inferensi, sementara data poisoning dapat meracuni model sejak tahap pelatihan, sehingga strategi pertahanan seperti adversarial training, validasi data, dan pemantauan drift model menjadi bagian penting dari siklus hidup sistem AI. Studi kasus shadow AI menunjukkan bahwa kesenjangan tata kelola AI membawa konsekuensi operasional dan finansial nyata, bukan sekadar risiko teoretis.

Daftar Pustaka

IBM Security & Ponemon Institute. (2025). Cost of a Data Breach Report 2025.
DataFence. (2025). Cost of a Data Breach 2025: IBM Report Analysis.
Kiteworks. (2025). How Shadow AI Costs Companies $670K Extra: IBM's 2025 Breach Report.
Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and Harnessing Adversarial Examples. International Conference on Learning Representations (ICLR).
Biggio, B., & Roli, F. (2018). Wild Patterns: Ten Years After the Rise of Adversarial Machine Learning. Pattern Recognition, 84, 317–331.

BAB 13

Federated Learning dan Kolaborasi Keamanan Antar-Organisasi

13.1 Pendahuluan

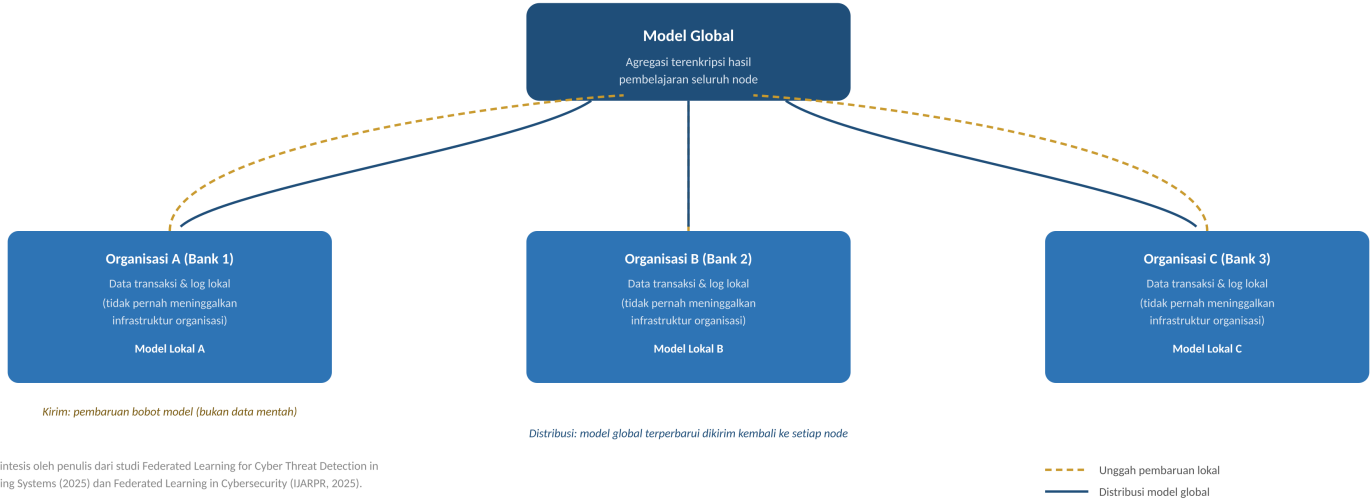
Bab 12 telah membahas ketegangan antara kebutuhan deteksi ancaman dan perlindungan privasi data. Bab ini membahas salah satu solusi teknis paling menjanjikan untuk ketegangan tersebut: federated learning, sebuah paradigma pembelajaran terdistribusi yang memungkinkan berbagai organisasi melatih model bersama tanpa harus saling berbagi data mentah sebuah kapabilitas yang sangat relevan untuk kolaborasi keamanan siber lintas-organisasi, mengingat sensitivitas data ancaman yang dimiliki masing-masing pihak.

13.2 Prinsip Kerja Federated Learning

Alih-alih mengumpulkan seluruh data pelatihan ke satu server pusat sebagaimana pada pembelajaran terpusat konvensional, federated learning membalik alur tersebut: model dikirim ke setiap node (organisasi), dilatih secara lokal menggunakan data milik organisasi tersebut, dan hanya pembaruan bobot model bukan data mentah yang dikirim kembali untuk diagregasi menjadi model global. Model global yang telah diperbarui kemudian didistribusikan kembali ke seluruh node, dan siklus ini berulang hingga model mencapai performa yang memadai.

Gambar 13.1 mengilustrasikan arsitektur ini dalam konteks kolaborasi tiga institusi perbankan yang ingin memperkuat deteksi ancaman siber bersama tanpa saling membuka data transaksi nasabah masing-masing.

Gambar 13.1 Arsitektur Federated Learning untuk Kolaborasi Keamanan Siber Antarorganisasi



Sumber: disintesis oleh penulis dari studi Federated Learning for Cyber Threat Detection in Digital Banking Systems (2025) dan Federated Learning in Cybersecurity (IJARPR, 2025).

Gambar 13.1 Arsitektur Federated Learning untuk Kolaborasi Keamanan Siber Antarorganisasi (disintesis oleh penulis)

Tabel 13.1 membandingkan pembelajaran terpusat konvensional dengan federated learning pada beberapa dimensi kunci yang relevan bagi keamanan siber.

Tabel 13.1 Perbandingan Pembelajaran Terpusat dan Federated Learning

Aspek	Pembelajaran Terpusat (Konvensional)	Federated Learning
Lokasi Data	Dikumpulkan dan disimpan di satu server pusat	Tetap berada di infrastruktur masing-masing organisasi
Privasi	Risiko tinggi karena data mentah berpindah dan terpusat	Privasi lebih terjaga karena hanya pembaruan model yang dipertukarkan (American Journal of AI Cyber Computing Management, 2025)
Kepatuhan Regulasi	Berpotensi melanggar regulasi lintas-yurisdiksi (mis. GDPR)	Lebih mudah memenuhi regulasi karena data tidak melewati batas yurisdiksi (International Journal of Advance Research Publication and Reviews, 2025)
Kualitas Model Kolaboratif	Terbatas pada data satu organisasi	Diperkaya oleh pola dari banyak organisasi tanpa berbagi data mentah (American Journal of AI Cyber Computing Management, 2025; Journal of Intelligent, Educational and Research, 2025)
Tantangan Utama	Skalabilitas infrastruktur terpusat	Heterogenitas data, latensi komunikasi, interpretabilitas (American Journal of AI Cyber Computing Management, 2025)

13.3 Studi Kasus: Federated Learning untuk Deteksi Ancaman Siber pada Perbankan Digital

Konteks dan Permasalahan

Sebuah penelitian tahun 2025 mengangkat permasalahan yang dihadapi institusi perbankan digital: setiap bank memiliki data transaksi dan pola ancaman yang berharga untuk melatih model deteksi yang lebih kuat, namun regulasi privasi data dan sensitivitas kompetitif membuat bank enggan berbagi data mentah mereka secara langsung dengan pihak lain, bahkan sesama institusi keuangan (American Journal of AI Cyber Computing Management, 2025).

Pendekatan

Studi ini menunjukkan bagaimana federated learning memungkinkan pelatihan model kolaboratif dan privacy-preserving pada berbagai node perbankan terdistribusi, memperkuat deteksi malware, phishing, dan anomali transaksional tanpa memerlukan sentralisasi data. Studi terkait lain yang meneliti kolaborasi lintas-yurisdiksi bahkan menunjukkan penerapan federated learning pada node yang tersebar secara geografis lintas negara dihubungkan melalui kanal komunikasi terenkripsi yang secara khusus dirancang untuk memitigasi ketegangan geopolitik sambil tetap mendukung deteksi kampanye advanced persistent threat (APT) secara kolektif (International Journal of Advance Research Publication and Reviews, 2025).

Hasil dan Pembelajaran

Penelitian ini menemukan bahwa federated learning berhasil meningkatkan privasi, kepatuhan regulasi, dan kualitas threat intelligence sekaligus, namun juga menghadapi tantangan nyata berupa heterogenitas

data antarbank (setiap bank memiliki karakteristik nasabah dan pola transaksi yang berbeda), latensi komunikasi dalam proses agregasi model, dan interpretabilitas hasil model gabungan (American Journal of AI Cyber Computing Management, 2025). Studi lintas-yurisdiksi turut mencatat bahwa pendekatan federated mendukung jaminan privasi yang dapat diaudit (auditable privacy guarantees), memungkinkan pengawasan hukum dan kebijakan dari otoritas keamanan siber nasional tanpa mengorbankan kedaulatan data masing-masing pihak (International Journal of Advance Research Publication and Reviews, 2025).

Catatan: Federated Learning sebagai Jembatan Kolaborasi dan Kedaulatan Data

- *Federated learning menawarkan solusi teknis konkret terhadap dilema yang dibahas pada Bab 12: bagaimana berkolaborasi memperkuat deteksi ancaman tanpa mengorbankan privasi dan kedaulatan data.*
- *Tantangan heterogenitas data antarorganisasi (non-IID data) tetap menjadi area riset aktif model global perlu dirancang agar tetap efektif meski pola data setiap node berbeda secara signifikan (American Journal of AI Cyber Computing Management, 2025).*
- *Skema secure aggregation, termasuk enkripsi homomorfik, semakin banyak diadopsi untuk memastikan bahkan pembaruan model yang dipertukarkan tidak dapat direkonstruksi kembali menjadi data asli (MDPI, 2025).*
- *Kolaborasi keamanan siber lintas-yurisdiksi melalui federated learning berpotensi menjadi model masa depan bagi berbagi ancaman (threat sharing) global, mengatasi hambatan kepercayaan geopolitik yang telah lama menghambat kolaborasi keamanan*

internasional (International Journal of Advance Research Publication and Reviews, 2025).

13.4 Rangkuman

Federated learning memungkinkan pelatihan model kolaboratif tanpa memerlukan sentralisasi data mentah, menjadikannya solusi teknis terhadap ketegangan privasi yang dibahas pada Bab 12. Prinsip kerja federated learning membalik alur pembelajaran konvensional: model dikirim ke data, bukan data dikirim ke model. Studi kasus perbankan digital menunjukkan bahwa federated learning secara nyata meningkatkan privasi dan kepatuhan regulasi, meski masih menghadapi tantangan heterogenitas data dan interpretabilitas. Kolaborasi lintas-yurisdiksi melalui federated learning berpotensi menjadi model masa depan bagi berbagi ancaman keamanan siber secara global.

Daftar Pustaka

American Journal of AI Cyber Computing Management. (2025). Federated Learning for Cyber Threat Detection in Digital Banking Systems.

International Journal of Advance Research Publication and Reviews. (2025). Federated Learning in Cybersecurity: Privacy-Preserving Collaborative Models for Threat Intelligence Across Geopolitically Sensitive Organizational Boundaries.

Journal of Intelligent, Educational and Research. (2025). Implementing Federated Learning in Threat Intelligence Sharing.

MDPI. (2025). Federated Learning-Driven Cybersecurity Framework for IoT Networks with Privacy Preserving and Real-Time Threat Detection Capabilities.

BAB 14

Arah Riset Masa Depan Menuju Sistem Pertahanan Siber yang Otonom

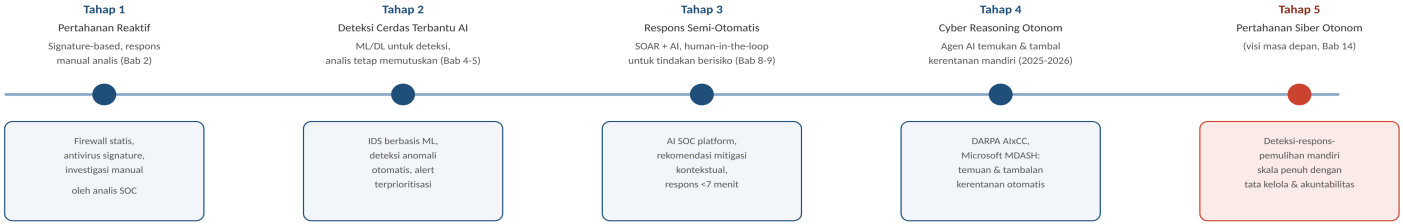
14.1 Pendahuluan

Sebagai penutup buku ini, bab terakhir membahas arah riset mutakhir dan proyeksi pengembangan sistem pertahanan siber otonom berbasis AI. Perjalanan pembahasan buku ini dari lanskap ancaman (Bab 1), keterbatasan pendekatan konvensional (Bab 2), taksonomi AI (Bab 3), fondasi teknis (Bagian II), hingga aplikasi dan tantangan (Bagian III dan IV) bermuara pada satu pertanyaan besar: seberapa jauh sistem pertahanan siber dapat, dan seharusnya, menjadi otonom?

14.2 Evolusi Menuju Pertahanan Siber Otonom

Gambar 14.1 mengilustrasikan evolusi bertahap dari pertahanan siber reaktif menuju visi pertahanan siber otonom sepenuhnya, memetakan bagaimana setiap tahap berkaitan dengan bab-bab yang telah dibahas di sepanjang buku ini.

Gambar 14.1 Evolusi Menuju Pertahanan Siber Otonom



Sumber: disintesis oleh penulis dari DARPA AIxCC Results (2025), Microsoft Security Blog (2026), dan AICA Reference Architecture (NATO, 2018).

Gambar 14.1 Evolusi Menuju Pertahanan Siber Otonom (disintesis oleh penulis)

Tabel 14.1 merangkum lima arah riset yang paling menonjol pada 2025–2026 dan diproyeksikan akan membentuk lanskap keamanan siber beberapa tahun mendatang.

Tabel 14.1 Arah Riset Masa Depan dalam AI untuk Keamanan Siber

Arah Riset	Deskripsi Singkat	Contoh Perkembangan Terkini
Cyber Reasoning System Otonom	Agen AI yang menemukan dan menambal kerentanan perangkat lunak secara mandiri	DARPA AI Cyber Challenge (AIxCC) 2025; Microsoft MDASH 2026 (DARPA, 2025; Cybersecurity Dive, 2025)
Agentic AI Multi-Peran	Beberapa agen AI berkolaborasi menjalankan peran berbeda dalam siklus keamanan	Torq HyperSOC, CORTEX (Bab 7-8) (Microsoft Security Blog, 2026)
Integrasi GNN + LLM + XAI	Kombinasi pemodelan relasional, pemahaman bahasa, dan transparansi	AboulEla & Kashef (2025) — Bab 6-7 (AboulEla & Kashef, 2025)
Pertahanan Siber Prediktif	Antisipasi ancaman sebelum tereksploitasi, bukan sekadar reaktif	Rapid7 2026 — model preemptive exposure reduction (Bab 1) (Rapid7 Labs, 2026)
Tata Kelola AI yang Matang	Standar dan regulasi yang mengimbangi kecepatan adopsi teknis	NIST Cyber AI Profile (Bab 10); AI Futures Steering Committee AS (Congress.gov / Library of Congress, 2026)

14.3 Studi Kasus: DARPA AI Cyber Challenge dan Cyber Reasoning System Otonom

Konteks dan Permasalahan

Defense Advanced Research Projects Agency (DARPA) menyelenggarakan AI Cyber Challenge (AIxCC), sebuah kompetisi dua tahun (2023–2025) yang menantang tim peneliti untuk membangun cyber reasoning system (CRS) yang sepenuhnya otonom mampu menemukan dan menambal kerentanan perangkat lunak sumber terbuka pada infrastruktur kritis tanpa campur tangan manusia, memanfaatkan kemajuan large language model yang telah dibahas pada Bab 7 (DARPA, 2025).

Pendekatan dan Hasil

Kompetisi final yang berlangsung Agustus 2025 melibatkan tujuh tim finalis yang menjalankan sistem CRS mereka secara otonom penuh selama sekitar 143 jam, menganalisis 53 proyek tantangan yang diturunkan dari perangkat lunak infrastruktur kritis nyata. Tim Atlanta keluar sebagai pemenang dengan hadiah 4 juta dolar AS, diikuti Trail of Bits dan Theori. DARPA mencatat bahwa kompetisi 2024–2025 ini berhasil mendemonstrasikan bahwa agen AI dapat menemukan dan memperbaiki kerentanan sumber terbuka dunia nyata, bahkan dalam beberapa kasus lebih cepat dibandingkan tim manusia (DARPA, 2025; Cybersecurity Dive, 2025).

Perkembangan lebih lanjut terjadi pada Mei 2026, ketika Microsoft mengumumkan MDASH (Multi-model Defensive Agentic Security at Hyperscale), sebuah sistem keamanan agentic multi-model yang dikembangkan sebagian oleh mantan anggota Tim Atlanta. Sistem ini berhasil membantu peneliti menemukan 16 kerentanan baru pada tumpukan jaringan dan otentikasi Windows, termasuk empat kerentanan kritis remote code execution pada komponen seperti kernel TCP/IP Windows dan layanan IKEv2 sebuah pencapaian yang menempatkan sistem

ini sebagai salah satu tolok ukur teratas dalam pengujian industri (Microsoft Security Blog, 2026).

Pembelajaran

Studi kasus ini menunjukkan bahwa visi "Tahap 4: Cyber Reasoning Otonom" pada Gambar 14.1 bukan lagi sekadar wacana futuristik, melainkan kapabilitas yang telah didemonstrasikan secara konkret pada infrastruktur perangkat lunak dunia nyata. Namun, penting dicatat bahwa kompetisi semacam AIxCC tetap beroperasi dalam lingkup terbatas (penemuan dan penambalan kerentanan perangkat lunak sumber terbuka), dan penerapan otonomi penuh pada seluruh siklus pertahanan siber organisasi termasuk keputusan respons berdampak tinggi yang dibahas pada Bab 8 dan Bab 12 masih memerlukan kerangka tata kelola dan akuntabilitas yang matang sebelum dapat diadopsi secara luas.

Catatan: Menyeimbangkan Kemajuan Otonomi dengan Tata Kelola

- *Kemajuan pesat menuju otonomi penuh (Bab 14) harus selalu diimbangi dengan prinsip tata kelola yang telah dibahas pada Bab 10 dan Bab 12 kecepatan teknis tanpa akuntabilitas yang memadai berisiko menciptakan "utang keamanan" baru, sebagaimana telah ditunjukkan oleh data IBM Cost of a Data Breach 2025.*
- *Riset seperti AIxCC dan MDASH menunjukkan bahwa arah masa depan keamanan siber bukan sekadar deteksi dan respons yang lebih cepat, melainkan pergeseran menuju pertahanan proaktif menemukan dan memperbaiki kerentanan sebelum sempat dieksploitasi.*
- *Komite pengarah kebijakan AI, seperti yang diamanatkan oleh National Defense Authorization Act FY2026 di Amerika Serikat,*

mencerminkan kesadaran yang berkembang bahwa kemajuan teknis otonomi perlu berjalan seiring dengan kerangka kebijakan yang matang.

- *Bagi pembaca buku ini mahasiswa, peneliti, maupun praktisi arah riset yang dibahas pada bab ini menawarkan peta jalan konkret untuk penelitian lanjutan, sejalan dengan semangat keberlanjutan riset yang telah dicontohkan pada studi kasus Bab 11.*

14.4 Penutup Buku

Perjalanan empat belas bab pada buku ini telah menelusuri lanskap ancaman siber yang terus berevolusi (Bagian I), fondasi teknis kecerdasan buatan dari machine learning klasik hingga graph neural network dan large language model (Bagian II), penerapan nyata pada sistem deteksi dan rekomendasi mitigasi yang terintegrasi dengan standar keamanan internasional (Bagian III), hingga tantangan etis dan arah masa depan menuju pertahanan siber otonom (Bagian IV). Benang merah yang menghubungkan seluruh pembahasan ini adalah satu prinsip sederhana namun mendasar: kecerdasan buatan menawarkan kapabilitas luar biasa untuk memperkuat ketahanan digital, tetapi kapabilitas tersebut hanya akan bernilai nyata jika dikembangkan dan diterapkan secara adaptif, transparan, dan bertanggung jawab.

14.5 Rangkuman

Evolusi pertahanan siber bergerak dari pendekatan reaktif menuju visi pertahanan otonom sepenuhnya, melalui tahapan deteksi cerdas, respons

semi-otomatis, dan cyber reasoning otonom. DARPA AI Cyber Challenge (AIxCC) 2025 dan Microsoft MDASH 2026 mendemonstrasikan bahwa agen AI otonom mampu menemukan dan menambal kerentanan perangkat lunak dunia nyata, termasuk kerentanan kritis pada infrastruktur luas seperti Windows. Kemajuan menuju otonomi penuh harus diimbangi dengan tata kelola yang matang, sejalan dengan prinsip yang dibahas pada Bab 10 dan Bab 12. Integrasi GNN, LLM, dan XAI diproyeksikan menjadi arah riset dominan, menyatukan kemampuan pemodelan relasional, pemahaman bahasa, dan transparansi keputusan.

Daftar Pustaka

DARPA. (2025). AI Cyber Challenge Marks Pivotal Inflection Point for Cyber Defense.

Cybersecurity Dive. (2025). DARPA Touts Value of AI-Powered Vulnerability Detection as It Announces Competition Winners.

Microsoft Security Blog. (2026). Defense at AI Speed: Microsoft's New Multi-Model Agentic Security System Tops Leading Industry Benchmark.

AboulEla, S. G., & Kashef, R. F. (2025). Leveraging Large Language Models, Graph Neural Networks, and Explainable AI for Next-Generation Network Intrusion Detection Systems.

Rapid7 Labs. (2026). 2026 Global Threat Landscape Report.

Congress.gov / Library of Congress. (2026). Agentic Artificial Intelligence and Cyberattacks (CRS Report IF13151).

BAB 15

Tantangan dan Praktik Industri dalam Adopsi AI untuk Keamanan Siber

15.1 Pendahuluan

Bab-bab sebelumnya pada buku ini telah membahas fondasi teknis kecerdasan buatan (Bagian II), penerapannya pada sistem deteksi dan rekomendasi mitigasi (Bagian III), serta tantangan etis dan arah masa depan (Bagian IV). Namun, terdapat kesenjangan yang perlu diakui secara eksplisit: kapabilitas teknis yang telah terbukti unggul secara akademik sebagaimana ditunjukkan berbagai studi kasus di sepanjang buku ini tidak secara otomatis diterjemahkan menjadi keberhasilan adopsi di lapangan. Bab ini secara khusus membahas fenomena dan tantangan nyata yang dihadapi industri dalam mengadopsi AI untuk keamanan siber, serta praktik-praktik konkret yang mulai diterapkan organisasi untuk mengatasinya.

15.2 Kesenjangan antara Kapabilitas Riset dan Kesiapan Operasional Industri

Survei enterprise AI adoption tahun 2026 mengungkap sebuah paradoks penting: meskipun kapabilitas teknis AI terus meningkat pesat, kesiapan

organisasi untuk mengadopsinya secara efektif tertinggal jauh di belakang. Laporan IBM tahun 2026 mencatat bahwa realitas adopsi AI saat ini ditandai oleh kapabilitas AI yang berkembang lebih cepat dibandingkan kapabilitas organisasi itu sendiri, dengan data yang terfragmentasi, tata kelola yang belum lengkap, kesenjangan talenta, dan kompleksitas sistem sebagai hambatan utama (IBM, 2026).

Survei terpisah yang dilakukan WRITER pada 2026 terhadap organisasi enterprise menemukan bahwa 79% organisasi menghadapi tantangan dalam mengadopsi AI meningkat dua digit dibandingkan 2025 dengan 54% eksekutif tingkat C-suite bahkan mengakui bahwa proses adopsi AI "memecah belah" organisasi mereka, meskipun 59% perusahaan telah menginvestasikan lebih dari satu juta dolar AS per tahun untuk teknologi AI (WRITER, 2026). Tabel 15.1 merangkum sejumlah indikator kunci dari berbagai survei industri terkini yang menggambarkan skala kesenjangan ini.

Tabel 15.1 Indikator Kesenjangan Adopsi AI dalam Industri Keamanan Siber (2026)

Indikator	Temuan Survei Industri 2026
Organisasi yang menghadapi tantangan dalam adopsi AI	79% (naik dua digit dari 2025)
Eksekutif C-suite yang menilai adopsi AI "memecah belah" organisasi mereka	54%
Organisasi yang berinvestasi >USD 1 juta/tahun pada teknologi AI	59%
Eksekutif yang meyakini organisasinya pernah mengalami kebocoran data akibat alat AI yang tidak disetujui (shadow AI)	67%
Organisasi dengan strategi keamanan AI tingkat lanjut	Hanya ±6%

Kesenjangan tenaga kerja keamanan siber global	4,8 juta pekerja
Analis SOC yang melaporkan kelelahan akibat volume alert (alert fatigue)	>70%

Yang secara khusus relevan bagi buku ini, temuan WRITER (2026) turut mengungkap bahwa 67% eksekutif meyakini organisasinya pernah mengalami kebocoran data akibat penggunaan alat AI yang tidak disetujui secara resmi sebuah angka yang menguatkan pembahasan risiko shadow AI pada Bab 12, menunjukkan bahwa fenomena ini bukan kekhawatiran teoretis, melainkan pengalaman nyata yang dilaporkan langsung oleh para pengambil keputusan di lapangan.

15.3 Tantangan Nyata: Kesenjangan Talenta, Fragmentasi Alat, dan Alert Fatigue

Tiga tantangan struktural yang telah disinggung secara singkat pada Bab 2 kembali muncul sebagai hambatan dominan dalam laporan-laporan industri terbaru, kini dengan data yang lebih tajam. Prediksi keamanan siber Palo Alto Networks untuk 2026 mencatat bahwa kesenjangan talenta keamanan siber global kini mencapai 4,8 juta pekerja sebuah "jurang permanen" (permanent chasm), bukan sekadar kesenjangan sementara dengan tim keamanan yang ada saat ini kewalahan oleh volume alert yang menyebabkan lebih dari 70% analis melaporkan alert fatigue (Palo Alto Networks, 2026).

Tantangan fragmentasi alat (tool fragmentation) turut ditegaskan sebagai akar masalah yang lebih mendasar daripada sekadar ketidaknyamanan operasional: alat keamanan yang terfragmentasi menciptakan silo data dan

titik buta (blind spot) yang membuat tata kelola yang dapat diverifikasi (verifiable governance) menjadi mustahil dicapai, menuntut pergeseran filosofi mendasar yang memandang risiko AI sebagai persoalan data, bukan sekadar persoalan alat (Palo Alto Networks, 2026). Yang mengejutkan, di tengah derasnya investasi dan adopsi AI, riset yang sama menemukan bahwa hanya sekitar 6% organisasi yang telah memiliki strategi keamanan AI tingkat lanjut, sementara Gartner memproyeksikan 40% aplikasi enterprise akan memiliki agen AI tugas-spesifik pada akhir 2026 kesenjangan tajam antara kecepatan adopsi teknis dan kematangan pengamanannya.

Survei CISO yang dilakukan KPMG pada 2026 terhadap 310 pemimpin keamanan di organisasi besar Amerika Serikat menambahkan dimensi lain: 74% pemimpin melaporkan peningkatan serangan siber, dan kompleksitas arsitektur keamanan yang terfragmentasi membuat banyak organisasi tetap berada dalam mode reaktif, meskipun tekanan untuk mempercepat adopsi AI generatif terus meningkat (KPMG, 2026).

15.4 Model Kematangan Organisasi dalam Adopsi AI untuk Keamanan Siber

Untuk membantu organisasi memahami posisi mereka dan merencanakan langkah selanjutnya, Gambar 15.1 menyajikan model kematangan lima tingkat yang disintesis penulis dari pola-pola yang teridentifikasi pada berbagai laporan industri 2026. Berbeda dengan evolusi teknis pertahanan siber otonom yang telah dibahas pada Bab 14 (Gambar 14.1), model ini berfokus pada kematangan organisasi tata kelola, konsolidasi alat, dan akuntabilitas yang menjadi prasyarat sebelum kapabilitas teknis canggih dapat memberikan nilai nyata.

Gambar 15.1 Model Kematangan Organisasi dalam Adopsi AI untuk Keamanan Siber

Survei industri 2026 menunjukkan mayoritas organisasi masih berada pada Level 1–2, dengan hanya sekitar 6% yang mencapai strategi keamanan AI tingkat lanjut

Sumber: disintesis oleh penulis dari IBM (2026), WRITER (2026), Palo Alto Networks (2026), dan World Economic Forum (2026).

Gambar 15.1 Model Kematangan Organisasi dalam Adopsi AI untuk Keamanan Siber (disintesis oleh penulis)

Temuan industri 2026 secara konsisten menunjukkan bahwa mayoritas organisasi masih berada pada Level 1 (Ad Hoc) atau Level 2 (Reaktif) tercermin dari statistik bahwa hanya sekitar 6% organisasi yang mencapai strategi keamanan AI tingkat lanjut (Palo Alto Networks, 2026). Pergerakan dari Level 2 menuju Level 3 (Terkelola) umumnya ditandai oleh konsolidasi platform keamanan yang sebelumnya terfragmentasi (Bab 8) dan mulai diterapkannya tinjauan berkala terhadap alat AI sebelum digunakan sebuah praktik yang menurut survei Global Cybersecurity Outlook 2026 dari World Economic Forum kini diterapkan oleh 40% organisasi, meningkat signifikan dari 24% yang sebelumnya hanya melakukan penilaian satu kali (World Economic Forum, 2026).

15.5 Praktik Industri: AI Security Posture Management (AI-SPM)

Salah satu praktik yang mulai menjadi standar industri pada 2026 adalah penerapan AI Security Posture Management (AI-SPM) dan Data Security Posture Management (DSPM) kelas alat yang secara khusus dirancang untuk memberikan visibilitas terhadap data dan model AI yang digunakan organisasi, mengatasi masalah "titik buta" yang dibahas pada Subbab 15.3. Prediksi industri menegaskan bahwa alat semacam ini akan menjadi kebutuhan mutlak (nonnegotiable) pada 2026, seiring meledaknya volume beban kerja dan data AI (Palo Alto Networks, 2026).

Yang secara khusus relevan bagi pembahasan data poisoning pada Subbab 12.6.2, laporan yang sama menggarisbawahi tantangan organisasi dalam mendeteksi manipulasi data yang tidak kasatmata: tim data yang memahami karakteristik data seringkali tidak terlatih mengenali manipulasi berbahaya, sementara tim keamanan yang memahami ancaman siber seringkali tidak memiliki visibilitas mendalam terhadap pipeline data

kesenjangan pemahaman inilah yang dimanfaatkan penyerang, karena data poisoning "tidak mendobrak pintu, melainkan menyelinap masuk menyamar sebagai data yang baik" (Palo Alto Networks, 2026).

15.6 MITRE ATLAS: Kerangka Kerja Standar Industri untuk Ancaman terhadap Sistem AI

Melengkapi pembahasan adversarial machine learning pada Bab 12, industri kini memiliki kerangka kerja standar untuk mengategorikan dan merespons ancaman terhadap sistem AI: MITRE ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems). Dimodelkan mengikuti kerangka MITRE ATT&CK yang telah banyak diadopsi dalam operasi keamanan siber konvensional, ATLAS menyediakan basis pengetahuan yang mendokumentasikan taktik dan teknik musuh yang secara spesifik menargetkan sistem AI dan machine learning, lengkap dengan puluhan studi kasus dunia nyata (Vectra AI, 2026).

Per pembaruan terakhir, ATLAS telah berkembang mencakup 16 taktik dan puluhan teknik serta sub-teknik, dengan pembaruan awal 2026 menambahkan cakupan khusus untuk ancaman terhadap AI agentic mencerminkan pergeseran fokus industri menuju keamanan sistem multi-agen yang telah disinggung pada Bab 8 dan Bab 14 (Vectra AI, 2026; Practical DevSecOps, 2026). Tabel 15.2 menyajikan contoh taktik ATLAS yang paling relevan dengan pembahasan adversarial machine learning pada buku ini.

Tabel 15.2 Contoh Taktik MITRE ATLAS dan Keterkaitannya dengan Buku Ini

Contoh Taktik MITRE ATLAS	Deskripsi Singkat	Keterkaitan dengan Buku Ini

Reconnaissance	Pengumpulan informasi tentang model target: arsitektur, data pelatihan, API inferensi	Tahap awal siklus serangan, sejalan dengan pola pengintaian pada Bab 1
ML Attack Staging	Persiapan serangan adversarial, termasuk pembuatan sampel perturbasi atau data racun	Berkorespondensi langsung dengan Subbab 12.6.1–12.6.2
ML Model Access	Memperoleh akses ke model melalui API publik, model curian, atau supply chain pihak ketiga	Relevan bagi organisasi yang mengandalkan model AI pihak ketiga (Bab 9-10)
Exfiltration	Mencuri model, data pelatihan, atau parameter melalui query API yang berulang (model extraction)	Ancaman terhadap kekayaan intelektual model GNN yang dibahas pada Bab 11
Impact	Dampak akhir: penurunan performa model, kegagalan deteksi, atau hilangnya kepercayaan pengguna	Konsekuensi langsung yang dibahas pada Subbab 12.6.3

MITRE ATLAS tidak dirancang untuk berdiri sendiri, melainkan melengkapi kerangka kerja tata kelola yang telah dibahas pada Bab 10: ATLAS berperan sebagai "sisi ancaman" yang mengidentifikasi dan mengkategorikan serangan, sementara NIST AI Risk Management Framework dan ISO/IEC 42001 berperan sebagai "sisi kontrol" yang mengatur bagaimana organisasi merespons dan mengelola risiko tersebut secara sistematis. Organisasi yang mengoperasionalkan ATLAS umumnya memulai dengan menyusun inventaris lengkap seluruh model dan agen AI yang mereka miliki sebuah

praktik dasar yang secara langsung sejalan dengan prinsip fungsi Identify pada NIST CSF 2.0 (Bab 10).

15.7 Studi Kasus: Kesenjangan antara Investasi dan Hasil Nyata

Konteks dan Permasalahan

Survei WRITER 2026 terhadap organisasi enterprise mengangkat permasalahan yang secara khusus relevan bagi bab ini: mengapa investasi besar pada teknologi AI (59% organisasi berinvestasi lebih dari satu juta dolar AS per tahun) tidak serta-merta menghasilkan transformasi organisasi yang diharapkan, dengan hanya 29% organisasi yang melaporkan ROI signifikan dari AI generatif (WRITER, 2026).

Temuan

Riset tersebut mengidentifikasi bahwa kegagalan transformasi bukan disebabkan oleh kurangnya talenta atau antusiasme terhadap AI, melainkan oleh ketiadaan sistem yang dirancang untuk menyebarkan praktik yang berhasil ke seluruh organisasi. Banyak organisasi memiliki "pengguna elite" (AI elite users) yang mencapai hasil luar biasa secara individual, namun tidak memiliki mekanisme untuk menyebarkan praktik tersebut secara organisasi-luas — individu berprestasi, namun tidak ada yang menghubungkan pencapaian tersebut dengan hasil bisnis yang terukur (WRITER, 2026).

Pembelajaran

Studi ini memberikan pembelajaran penting yang melengkapi keseluruhan buku ini: kapabilitas teknis AI yang unggul baik itu model GNN pada Bab 6 dan 11, sistem rekomendasi mitigasi pada Bab 9, maupun kerangka tata

kelola pada Bab 10 hanya akan memberikan nilai nyata bagi organisasi jika disertai investasi setara pada sisi organisasional: struktur akuntabilitas yang jelas, kerangka tata kelola yang matang (Bab 10, Bab 12), dan mekanisme sistematis untuk menyebarkan praktik yang terbukti berhasil ke seluruh unit organisasi. Kesenjangan antara kapabilitas teknis dan kesiapan organisasi inilah yang menjadi benang merah seluruh pembahasan bab ini.

Catatan: Pelajaran Praktis bagi Organisasi yang Mengadopsi AI Keamanan Siber

- *Mulailah dari tata kelola, bukan teknologi kebijakan penggunaan AI yang jelas dan disosialisasikan luas terbukti lebih menentukan keberhasilan adopsi dibandingkan kecanggihan model semata (Bab 10, Bab 12).*
- *Konsolidasi alat keamanan yang terfragmentasi menjadi prasyarat penting sebelum kapabilitas AI tingkat lanjut dapat memberikan visibilitas dan nilai yang optimal.*
- *Adopsi kerangka kerja standar seperti MITRE ATLAS untuk mengategorikan ancaman terhadap sistem AI, dipadukan dengan NIST AI RMF dan ISO/IEC 42001 sebagai kerangka kontrol (Bab 10).*
- *Investasi pada AI-SPM/DSPM dan tinjauan berkala terhadap alat AI (bukan sekadar penilaian satu kali di awal) merupakan praktik yang semakin menjadi standar industri pada 2026.*
- *Keberhasilan transformasi AI menuntut mekanisme organisasional untuk menyebarkan praktik terbaik secara luas, bukan hanya bergantung pada segelintir "pengguna elite" yang berprestasi secara individual.*

15.8 Rangkuman

Kesenjangan antara kapabilitas teknis AI yang terus berkembang pesat dan kesiapan operasional organisasi merupakan tantangan dominan yang dihadapi industri pada 2026, dengan 79% organisasi melaporkan kesulitan adopsi meskipun investasi besar telah dilakukan. Tiga tantangan struktural kesenjangan talenta (4,8 juta pekerja secara global), fragmentasi alat keamanan yang menciptakan titik buta, dan alert fatigue yang melanda lebih dari 70% analis terus menjadi hambatan utama, diperparah oleh kenyataan bahwa hanya sekitar 6% organisasi memiliki strategi keamanan AI tingkat lanjut. Model kematangan lima tingkat yang disajikan pada bab ini memetakan perjalanan organisasi dari adopsi ad hoc tanpa tata kelola menuju integrasi AI yang teroptimasi dan akuntabel, dengan mayoritas organisasi saat ini masih berada pada tingkat awal. Praktik industri yang mulai berkembang AI Security Posture Management, tinjauan berkala terhadap alat AI, dan adopsi kerangka kerja standar seperti MITRE ATLAS yang melengkapi NIST AI RMF dan ISO/IEC 42001 menunjukkan arah konkret bagi organisasi untuk menutup kesenjangan tersebut. Studi kasus mengenai disparitas antara investasi dan hasil nyata menggarisbawahi pelajaran inti bab ini: keunggulan teknis semata tidak cukup, dan keberhasilan adopsi AI untuk keamanan siber pada akhirnya ditentukan oleh kematangan organisasional yang menyertainya.

Daftar Pustaka

IBM. (2026). The Biggest AI Adoption Challenges for 2026.

WRITER. (2026). Enterprise AI Adoption in 2026: Why 79% Face Challenges Despite High Investment.

Palo Alto Networks. (2026). 2026 Cybersecurity Predictions.

KPMG. (2026). 2026 Cybersecurity & Technology Risk Survey: The CISO's Evolving Role.

World Economic Forum. (2026). Global Cybersecurity Outlook 2026: The Trends Reshaping Cybersecurity.
Vectra AI. (2026). MITRE ATLAS: AI Security Framework with 16 Tactics and 84 Techniques.
Practical DevSecOps. (2026). MITRE ATLAS Framework 2026: Guide to Securing AI Systems.
MarketScale Newsroom. (2026). Enterprise AI Moves from Pilot to Production in 2026, but Gaps in Governance and Talent Persist.

LAMPIRAN A — Glosarium Istilah Kecerdasan Buatan dan Keamanan Siber

Adversarial Attack
Masukan yang secara sengaja dirancang untuk mengelabui model machine learning agar salah klasifikasi.

Agentic AI
Sistem AI yang mampu merencanakan dan menjalankan serangkaian tindakan secara otonom untuk mencapai tujuan tertentu.

Anomaly Detection
Teknik mengenali pola data yang menyimpang secara signifikan dari perilaku normal.

Attack Surface
Keseluruhan titik dalam sistem yang berpotensi dieksploitasi pihak tidak berwenang.

Autoencoder
Arsitektur neural network yang belajar merekonstruksi data normal; digunakan untuk deteksi anomali.

CIA Triad
Tiga pilar keamanan informasi: Confidentiality (kerahasiaan), Integrity (integritas), Availability (ketersediaan).

Deep Learning
Subbidang machine learning yang menggunakan neural network berlapis untuk mempelajari representasi fitur secara otomatis.

Deepfake
Konten audio/video sintetis yang dihasilkan AI untuk meniru identitas seseorang secara meyakinkan.

Federated Learning

Paradigma pembelajaran terdistribusi yang melatih model bersama tanpa memusatkan data mentah.

GNN (Graph Neural Network)

Arsitektur neural network yang dirancang untuk mempelajari representasi dari data berstruktur graf.

Human-in-the-Loop

Model operasi yang mempertahankan keterlibatan manusia dalam keputusan AI, khususnya untuk tindakan berdampak tinggi.

IDS/IPS

Intrusion Detection System / Intrusion Prevention System — sistem deteksi dan/atau pencegahan intrusi jaringan.

LLM (Large Language Model)

Model bahasa berskala besar berbasis arsitektur transformer yang mampu memahami dan menghasilkan teks.

Machine Learning

Cabang kecerdasan buatan yang memungkinkan sistem belajar dari data tanpa diprogram secara eksplisit.

NIST CSF

NIST Cybersecurity Framework — kerangka kerja manajemen risiko siber yang dikembangkan oleh National Institute of Standards and Technology, AS.

Ransomware-as-a-Service (RaaS)

Model bisnis kejahatan siber di mana operator ransomware menyewakan alat serangan kepada afiliasi.

Shadow AI

Penggunaan alat AI oleh individu dalam organisasi tanpa persetujuan atau pengawasan resmi tim keamanan/TI.

Signature-based Detection

Metode deteksi ancaman dengan mencocokkan pola terhadap basis data tanda tangan ancaman yang telah dikenal.

XAI (Explainable AI)

Pendekatan yang membuat keputusan model AI dapat dipahami dan dijelaskan kepada manusia.

Zero-Day Attack

Serangan yang mengeksploitasi kerentanan yang belum diketahui atau belum ditambal oleh vendor perangkat lunak.

LAMPIRAN B — Daftar Dataset Publik untuk Riset Deteksi dan Mitigasi Serangan Siber

Lampiran ini merangkum dataset publik yang telah dibahas di sepanjang buku ini, sebagai referensi cepat bagi pembaca yang ingin memulai riset atau eksperimen lanjutan.

Tabel B.1 Dataset Publik untuk Riset Keamanan Siber Berbasis AI

Nama Dataset	Domain	Kegunaan Utama
NSL-KDD	Deteksi intrusi jaringan	Benchmark klasik untuk klasifikasi trafik normal vs. serangan (Bab 4)
CICIDS2017 / CSE-CIC-IDS2018	Deteksi intrusi jaringan	Trafik jaringan realistis dengan skenario serangan modern (Bab 4)
UNSW-NB15	Deteksi intrusi jaringan	Kombinasi trafik normal dan sintetik dengan pola serangan kontemporer (Bab 4)
CIC-MalMem-2022	Deteksi malware berbasis memori	Deteksi malware tersembunyi (fileless malware) (Bab 4-5)
Credit Card Fraud Dataset / NeurIPS 2022 Bank Account Fraud	Deteksi fraud finansial	Benchmark deteksi anomali transaksi (Bab 4)
Elliptic Dataset (Bitcoin)	Deteksi fraud pada graf transaksi	Benchmark graph-based fraud detection (Bab 6)

Bagaimana AI Mengubah Cara Kita Melawan Ancaman Siber

Konsep, Model dan Implementasi

Bagaimana AI Mengubah Cara Kita Melawan Ancaman Siber hadir sebagai jawaban atas kesenjangan langka namun krusial: minimnya literatur berbahasa Indonesia yang membahas irisan kecerdasan buatan dan keamanan siber secara utuh mulai dari fondasi matematis machine learning, deep learning, dan graph neural network, hingga implementasi nyata dalam bentuk sistem deteksi dan rekomendasi mitigasi yang teruji pada data serangan sungguhan. Buku ini penting karena tidak berhenti di teori: setiap konsep dijelaskan dengan pseudocode dan studi kasus konkret dari platform AI SOC industri, kerangka NIST/ISO, hingga purwarupa GNN yang diuji langsung pada infrastruktur honeypot sekaligus jujur mengakui keterbatasan dan risiko adopsi AI itu sendiri, menjadikannya bacaan yang relevan bagi mahasiswa, peneliti, maupun praktisi yang ingin memahami bagaimana pertahanan digital Indonesia dapat diperkuat di tengah ancaman siber yang terus berevolusi.

Penerbit
REDTECH PUTRA BENUA
Perumahan Graha Jatikuwung Indah D04
Jatikuwung, Gondangrejo, Karanganyar
Phone. 08161 355 055
Webpage : www.redtechidn.org

ISBN 978-634-05-3679-9 (PDF)
9 786340 536799