Kombinasi Agregasi Data Jaringan dan Isi Email untuk Deteksi Anomali pada Serangan *Spear Phising*

I Made Wisma Fajaryasa[1)], Dandy Pramana Hostiadi[2)], Roy Rudolf Huizen[3)]
Magister Sistem Informasi[1)2)3)]
Institut Teknologi dan Bisnis STIKOM Bali
Denpasar, Indonesia
e-mail: 232012009@stikom-bali.ac.id[1)], dandy@stikom-bali.ac.id[2)], roy@stikom-bali.ac.id[3)]

Abstrak

Perkembangan era digitalisasi telah mengalami perkembangan yang sangat signifikan, dengan demikian peningkatan pengguna media layanan yang menggunakan sebagai sarana komunikasi dan layanan internet cenderung mengalami peningkatan namun diimbangi dengan kenaikan potensi kejahatan siber. Salah satunya adalah serangan spear phising merupakan ancaman siber yang semakin canggih dengan memanfaatkan informasi spesifik target untuk menipu dan mencuri data sensitif. Deteksi serangan ini menjadi tantangan karena metode konvensional sering kali gagal mengidentifikasi anomali dengan akurat. Penelitian ini mengusulkan pendekatan kombinasi agregasi data jaringan dan isi email sebanyak 11.050 data dengan 17 fitur karakteristik email dan 13 fitur karakteristik network activity untuk meningkatkan deteksi anomali pada serangan spear phising. Metode ini juga melakukan analisis komparatif, di mana hasil evaluasi menunjukkan nilai rata-rata sebesar 0.984 untuk agregasi data antara email dan jaringan, sementara data email memperoleh nilai rata-rata 0.941, dan data jaringan memperoleh nilai rata-rata 0.755. Hasil eksperimen menunjukkan bahwa pendekatan ini mampu meningkatkan tingkat anomali terhadap serangan spear phising dengan demikian, penelitian ini memberikan kontribusi dalam pengembangan sistem keamanan siber yang lebih adaptif terhadap ancaman phishing modern.

Kata kunci: *Spear phising, Network Activity, Body Email, Anomali Detection, Machine Learning.*

1. Pendahuluan

Teknologi informasi mampu mengubah realitas ekonomi, budaya, politik dan hukum. Seiring berkembangnya teknologi informasi mampu memberikan dampak positif bagi banyak orang namun hal ini juga menyebabkan munculnya kejahatan-kejahatan baru yang disebut dengan kejahatan dunia maya baru melalui jaringan internet[1]. Dimana terdapat beberapa orang yang memanfaatkan celah keamanan pada teknologi informasi pada jaringan internet sebagai sarana untuk melakukan kejahatan yang selanjutnya dikenal dengan cybercrime[2]. Cybercrime merupakan perbuatan melawan hukum yang dilakukan dengan menggunakan jaringan komputer atau internet yang berbasis pada kecanggihan teknologi komputer dan telekomunikasi untuk memperoleh keuntungan dengan merugikan pihak lain[3]. Salah satu bentuk cybercrime yang paling umum dan merugikan adalah phishing. Phishing adalah upaya penipuan melalui teknik manipulasi psikologis untuk memperoleh informasi sensitif, seperti kata sandi, nomor kartu kredit, atau data pribadi, dengan menyamar sebagai entitas yang tepercaya[4].

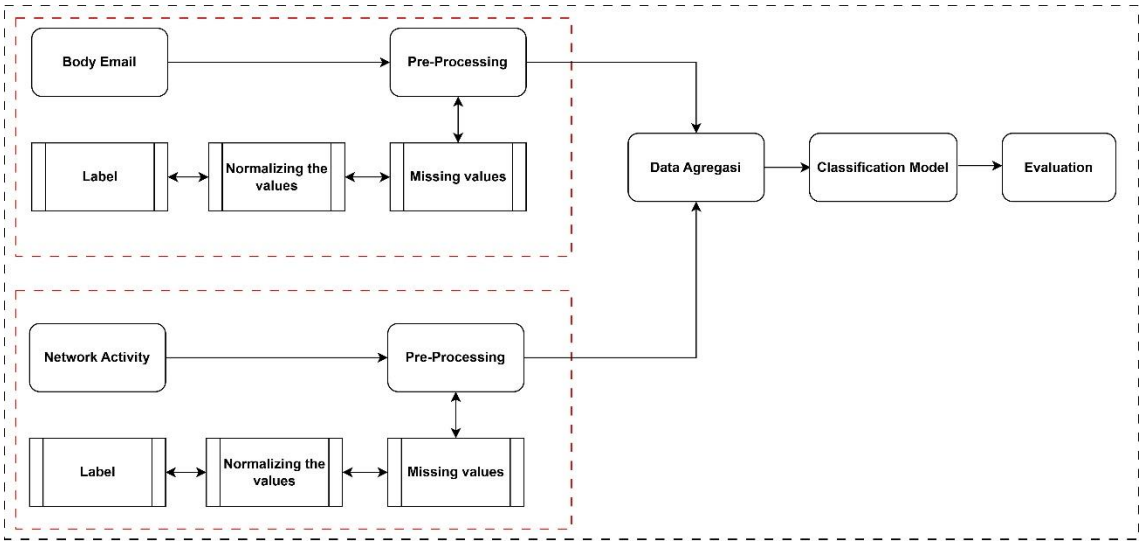
Serangan ini sering kali bertujuan untuk mencuri informasi sensitif seperti data pribadi, kredensial login, atau bahkan akses ke sistem internal perusahaan[5]. Phishing pada email merupakan tindakan penipuan yang dilakukan dengan cara mengirimkan email palsu yang seolah-olah berasal dari sumber terpercaya untuk mendapatkan informasi sensitif dari penerima, seperti kata sandi, nomor kartu kredit, atau data pribadi lainnya[6]. Oleh karenanya, email dapat dengan mudah dimanipulasi agar terlihat seperti berasal dari sumber terpercaya, seperti bank, perusahaan teknologi, atau instansi pemerintah. Hal ini membuat penerima lebih mudah terperdaya. Faktor di atas membuat email menjadi salah satu medium favorit bagi pelaku phishing untuk melancarkan aksinya[7]. Kesadaran pengguna dan penerapan langkah-langkah keamanan menjadi kunci untuk menghindari serangan ini.

Deteksi serangan spear phishing menjadi tantangan tersendiri karena teknik yang digunakan semakin canggih dan seringkali mampu mengelabui sistem keamanan tradisional. Pendekatan konvensional yang hanya berfokus pada analisis isi email atau aktivitas jaringan secara terpisah terbukti kurang efektif dalam mengidentifikasi anomali yang kompleks[8]. Oleh karena itu, diperlukan metode yang lebih komprehensif dengan menggabungkan analisis data jaringan dan konten email untuk meningkatkan akurasi

deteksi. Penelitian ini mengusulkan kombinasi Agregasi data jaringan dan isi email sebagai pendekatan baru dalam mendeteksi anomali pada serangan spear phishing. Dengan menganalisis 11.050 data yang mencakup 17 fitur karakteristik email dan 13 fitur aktivitas jaringan, diharapkan metode ini mampu mengidentifikasi pola-pola mencurigakan yang tidak terdeteksi oleh sistem konvensional[9]. Hasil eksperimen menunjukkan bahwa pendekatan ini berhasil meningkatkan tingkat deteksi anomali, dengan nilai rata-rata akurasi sebesar 0,974 untuk kombinasi data, dibandingkan dengan analisis terpisah yang menghasilkan nilai rata-rata lebih rendah. Temuan ini memberikan kontribusi penting dalam pengembangan sistem keamanan siber yang lebih adaptif dan responsif terhadap ancaman phishing modern. Implementasi pendekatan kombinasi ini diharapkan dapat memperkuat pertahanan terhadap serangan spear phishing dan meminimalkan risiko kebocoran data sensitif di berbagai sektor.

2. Metode Penelitian

Pada Penelitian yang diusulkan ini dengan melakukan perkenalan model untuk mendapatkan hasil yang lebih baik dalam akurasi dengan menggunakan klasifikasi individu dengan machine learning yaitu *Deceision Tree*. Arsitektur model sistem akan dijelaskan pada bagian ini dan juga akan disampaikan teknik Data Agregasi dan algoritma klasifikasi. Secara umum, arsitektur model sistem yang diusulkan ditampilkan pada Gambar 1.



Gambar 1. Arsitektur Model

2.1 Arsitektur Model

Langkah awal pada bagian ini adalah melakukan preprocessing untuk masing-masing dataset *body email* dan *network activity*, setelah itu akan dilakukan kombinasi dua data manjadi satu yaitu dengan melakukan kombinasi dari data *Body email* dan data *Network Activity*, dimana dari kedua data tersebut akan menentukan dua kelas yaitu sebagai *malicious* dan normal dengan mengimplementasikan klasifikasi *machine learning Decission Tree* sebagai model deteksi dari anomali serangan *spear phising*. Hal ini terdiri dari beberapa tahapan: *Pre-processing Body Email* Dataset dan *Network Activity* Dataset, Data Agregasi, *Classification Model*, dan *Evaluation.*

a) Pre-Processing

Dataset yang digunakan terdiri dari dua sumber utama: (1) Body Email Dataset yang mencakup konten email dan label terkait, serta (2) Network Activity Dataset yang berisi rekaman aktivitas jaringan yang berpotensi mencerminkan serangan. Pada tahap *pre-processing* data, dilakukan serangkaian langkah untuk meningkatkan kualitas dataset sebelum digunakan dalam analisis dan pemodelan. Proses dimulai dengan pembersihan data, di mana duplikasi dan noise dihapus guna meningkatkan akurasi serta menghindari bias dalam model. Selanjutnya, dilakukan transformasi fitur yang mencakup normalisasi untuk menyelaraskan skala fitur numerik, standarisasi untuk memastikan distribusi data yang lebih stabil guna meningkatkan representasi data. Setelah data mengalami transformasi. Dengan proses ini, dataset yang dihasilkan memiliki kualitas tinggi dan lebih siap digunakan untuk mendeteksi anomali pada serangan *spear phising.*

b) Data Agregasi

Proses Data agregasi dilakukan dengan menggabungkan data dari *Body Email* dan *Network Activity* menjadi satu dataset terpadu untuk analisis lebih mendalam. Integrasi ini memungkinkan pengolahan informasi dari berbagai sumber secara bersamaan, sehingga memberikan konteks yang lebih lengkap dalam mendeteksi pola serangan *spear phising.* Tujuan dari agregasi ini adalah menggabungkan informasi dari email dan aktivitas jaringan sehingga dapat memberikan pemahaman yang lebih komprehensif terhadap pola anomali yang mungkin muncul dalam serangan spear-phishing Dengan menyatukan data tekstual dari *body email* dan data aktivitas jaringan, proses ini memperkuat analisis dengan menyediakan wawasan yang lebih komprehensif, meningkatkan akurasi dalam identifikasi serangan[10].

c) *Classification Model*

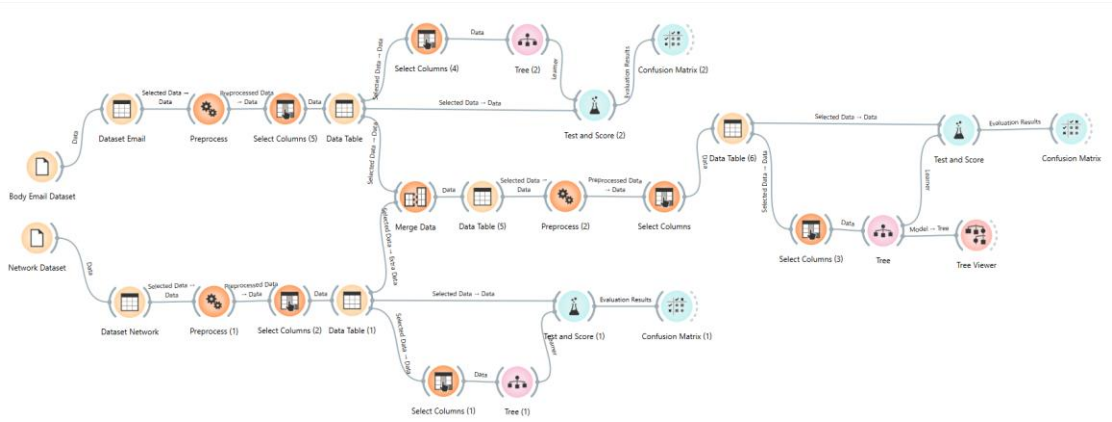
Data yang telah digabungkan digunakan sebagai input untuk model klasifikasi. Model ini bertujuan untuk mengidentifikasi apakah suatu instance termasuk dalam kategori spear-phishing atau tidak. Dalam penelitian ini evaluasi kinerja model dilakukan dengan menggunakan *confusion matrix*. *Confusion matrix* memberikan gambaran visual tentang performa model, termasuk jumlah *true positives*, *true negatives*, *false positives*, dan *false negatives*. *Confusion matrix* yang digunakan adalah akurasi, presisi, recall, dan F1-score.

d) Evaluation

Hasil dari evaluasi model dianalisis untuk menentukan seberapa baik model dapat mendeteksi anomali terhadap serangan *spear phising.* Analisis ini mencakup diskusi tentang kekuatan dan kelemahan model, serta potensi perbaikan yang dapat dilakukan di masa mendatang[11]. Tahapan akhir dari penelitian adalah menarik kesimpulan berdasarkan hasil analisis. Kesimpulan mencakup ringkasan temuan utama, implikasi dari hasil penelitian, serta rekomendasi untuk penelitian lebih lanjut dan implementasi di bidang deteksi serangan *spear phising.*

3. Hasil dan Pembahasan

Bagian ini menyajikan hasil dan pembahasan dari penelitian yang menggunakan data UCI Repository sebagai studi kasus. Pada penelitian ini mengimplementasikan dan mengevaluasi metode yang diusulkan dengan menggunakan *software Orange Data Mining* pada Laptop CPU : Intel Core i5 12450H dan RAM : 8 Gb , SSD : 500 Gb. Seperti pada Gambar 2 yang berisi tentang proses klasifikasi pada *software Orange Data Mining*.



Gambar 2 Proses Klasifikasi pada *Software Orange Data Mining*

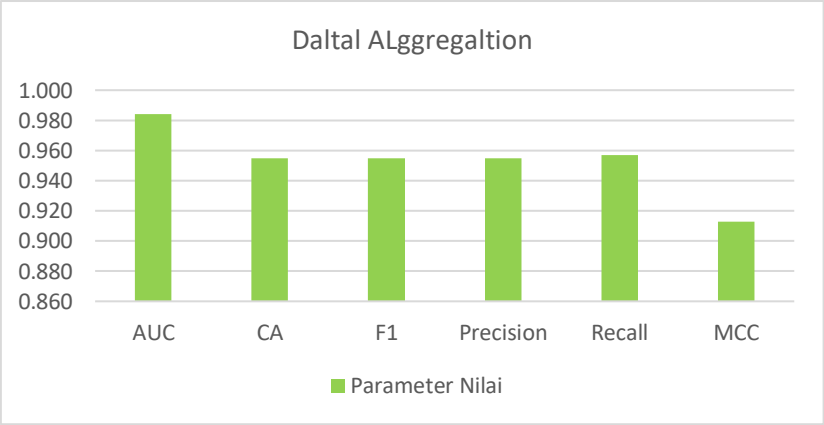
Data UCI Repository memiliki data sebanyak 11.506 data dan terbagi kedalam dua file CSV dengan 17 fitur untuk *body email* dataset dan 13 fitur untuk *network activity* dataset hasil dari fitur tersebut akan memiliki 1 kelas yang menunjukan nilai malicious dan normal. Pada masing-masing fitur pada *body email* dan *network* memiliki karakteristik. Dimana, 17 fitur *body email* terhadap serangan *spear phising* yaitu having_At_Symbol, double_slash_redirecting, Prefix_Suffix, having_Sub_Domain, SSLfinal_State, Favicon, Request_URL, URL_of_Anchor, Links_in_tags, SFH, Submitting_to_email, on_mouseover, RightClick, popUpWidnow, Google_Index, Links_pointing_to_page, dan Statistical_report dan 13 fitur *network activity* terhadap serangan *spear phising* yaitu having_IP_Address, URL_Length,

Shortining_Service, Domain_registeration_length, port, Abnormal_URL, Redirect, Iframe, age_of_domain, DNSRecord, web_traffic, dan Page_Rank. Seperti yang ditunjuk pada Tabel 1.

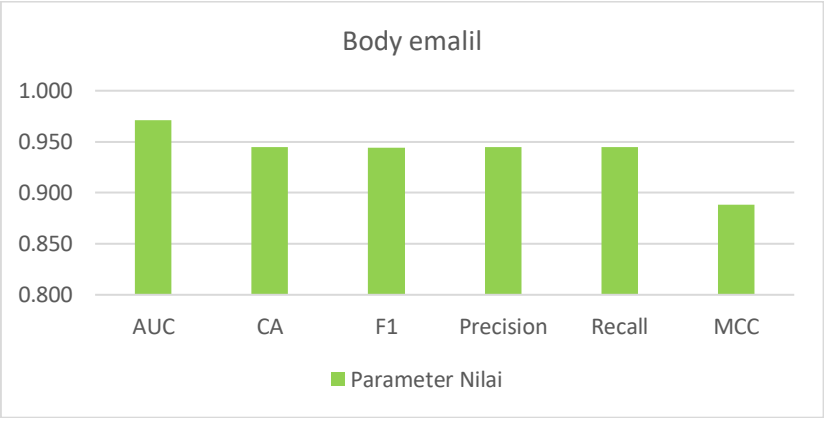
Table 1 Dataset *Spear Phising*

Dataset	Jumlah Data	Jumlah Fitur
Body Email	11.055	17
Network Activity	11.055	13

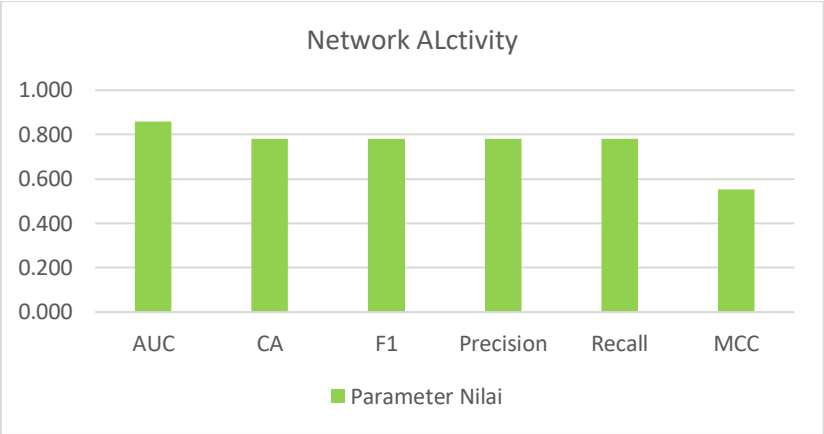
Proses awal pengolahan data adalah data transformasi, yaitu perubahan data kategorikal menjadi numerik. Hasil dari kombinasi data antara data *body email* dan *network activity* akan menjadi 30 fitur sebagai parameter untuk melakukan deteksi terkait dengan anomali yang terjadi pada serangan *spear phising*, kemudian dilakukan proses data normalisasi, yaitu membentuk skala data menjadi rentang nilai 0-1. Setelah itu, data di bagi menjadi dua yaitu data latih yang digunakan dalam pelatihan model dan data uji yang digunakan untuk menguji model klasifikasi. Dalam penelitian ini digunakan komposisi data yaitu 70% untuk pelatihan dan 30% untuk pengujian. Hasil pemodelan menunjukan bahwa model Decesion tree memiliki rata-rata performa diatas 0.95 untuk kombinasi data antara *body email* dan *network activity*, sedangkan menggunakan *body email* dan menggunakan *network activity* memiliki rata-rata performa dibawahnya terhadap prediksi anomali terkait dengan serangan *spear phising*. Hasil performa model ditunjukan pada gambar berikut.



Gambar 3 Hasil Deteksi Data Aggregation dengan metode Decission tree



Gambar 4 Hasil Deteksi data Body Email dengan metode Decission tree

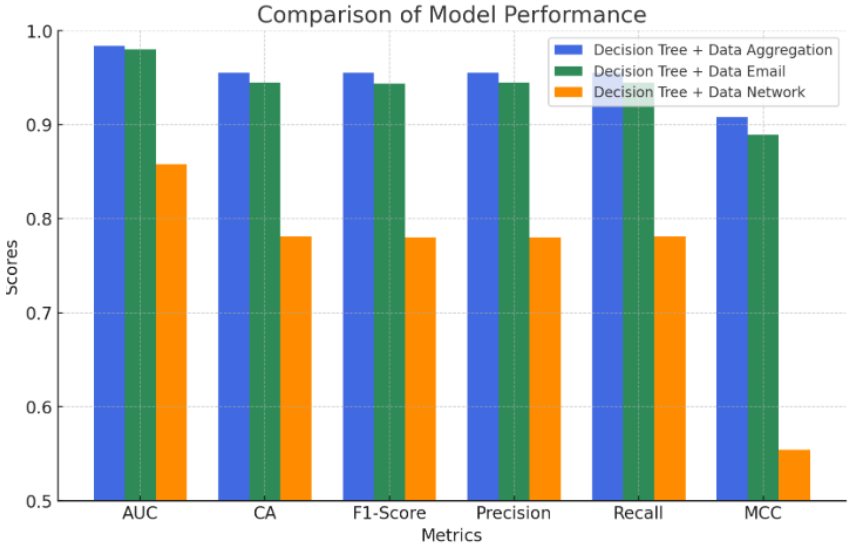


Gambar 5 Hasil Deteksi data Network Activity dengan metode Decission tree

Hasil dari masing-masing data set dan kombinasi dataset dapat menghasilkan perbandingan atau komparasi hasil terhadap masing-masing karakteristik dataset, dengan pengolahan atau pre-proccess data yang sama maka dapat dilihat dari hasil akurasi keseluruhan kinerja klasifikasi dinyatakan dalam hal AUC, CA, F1-Score, Precision, Recall, dan MCC. Perbandingan menunjukan bahwa usulan teknik kombinasi dataset antara dataset email dan dataset network memiliki nilai yang lebih baik dari parameter AUC, CA, F1-Score, Precision, Recall dan MCC. Hasil komparasi dapat dilihat pada table.

Table 2 Komparasi model deteksi anomali *spear phising*

Model	AUC	CA	F1-Score	Precision	Recall	MCC
Deceision Tree + Data Aggregation	0.984	0.955	0.955	0.955	0.955	0.908
Deceision Tree + Data Email	0.980	0.945	0.944	0.945	0.945	0.889
Deceision Tree + Data Network	0.858	0.781	0.780	0.780	0.781	0.554



Gambar 6 *Comparison of Model Performance*

4. Kesimpulan

Penelitian ini mengusulkan pendekatan kombinasi agregasi data jaringan dan isi email untuk meningkatkan deteksi anomali pada serangan spear phishing. Dengan menganalisis 11.050 data yang mencakup 17 fitur karakteristik email dan 13 fitur aktivitas jaringan, metode ini terbukti lebih efektif dibandingkan analisis konvensional yang hanya berfokus pada satu jenis data. Hasil evaluasi menunjukkan bahwa kombinasi data memiliki akurasi rata-rata 0,984, lebih tinggi dibandingkan dengan analisis data

email 0,941 dan data jaringan 0,755. Meskipun pendekatan ini masih memiliki keterbatasan, kombinasi dataset terbukti berpengaruh terhadap peningkatan performa deteksi anomali. Penelitian lanjutan diharapkan dapat mengeksplorasi metode klasifikasi lain dan menambahkan fitur baru untuk meningkatkan akurasi sistem deteksi serangan *spear phishing*.

Daftar Pustaka

[1] J. Wang, T. Herath, R. Chen, A. Vishwanath, and H. R. Rao, "Research article phishing susceptibility: An investigation into the processing of a targeted spear phishing email," *IEEE Trans. Prof. Commun.*, vol. 55, no. 4, pp. 345–362, 2012, doi: 10.1109/TPC.2012.2208392.

[2] M. Fadhlurrohman, A. Muliawati, and B. Hananto, "Analisis Kinerja Intrusion Detection System pada Deteksi Anomali dengan Metode Decision Tree Terhadap Serangan Siber," *J. Ilmu Komput. dan Agri-Informatika*, vol. 8, no. 2, pp. 90–94, 2021, doi: 10.29244/jika.8.2.90-94.

[3] D. Bringhenti *et al.*, "Automatic, verifiable and optimized policy-based security enforcement for SDN-aware IoT networks," *Comput. Networks*, vol. 213, no. January, p. 109123, 2022, doi: 10.1016/j.comnet.2022.109123.

[4] A. Wibowo Noor Fikri *et al.*, "Analisis Keamanan Sistem Operasi dalam Menghadapi Ancaman Phishing dalam Layanan Online Banking," *J. Ilmu Multidisplin*, vol. 2, no. 1, pp. 84–91, 2023, doi: 10.38035/jim.v2i1.228.

[5] G. S. W. Prabuningrat, D. P. Hostiadi, and N. L. P. Srinadi, "Klasifikasi Deteksi Anomali Menggunakan Metode Machine Learning," *Pros. Semin. Has. Penelit. Inform. dan Komput.*, vol. 1, no. 2, pp. 845–850, 2024.

[6] G. Wibisono, R. A. G. Gultom, and T. Mantoro, "Strategi Peningkatan Kapabilitas Satuan Siber Dispamsanau Melalui Pemanfaatan Artificial Intelligence Pada Keamanan Siber Berdasarkan National Institute of Standards and Technology Cybersecurity Framework Version 1.1," *J. Rev. Pendidik. dan Pengajaran*, vol. 7, no. 1, pp. 968–975, 2024.

[7] D. A. Daniswara, A. Budiyono, A. Almaarif, and S. Kom, "Analisis Deteksi Malicious Activity Menggunakan Metode Analisis Malware Dinamis Berbasis Anomali Detection Analysis of Malicious Activity Using Anomaly-Based Dynamic Malware Analysis Method," *Anal. Deteksi Malicious Act. Menggunakan Metod. Anal. Malware Din. Berbas. Anomali Detect. Anal. Malicious Act. Using Anomaly-Based Dyn. Malware Anal. Method*, vol. 6, no. 2, pp. 3–8, 2019, [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/10635

[8] G. M. G. Bororing, "Pengembangan Algoritma Machine Learning Untuk Mendeteksi Anomali Dalam Jaringan Komputer," *J. Rev. Pendidik. dan ...*, vol. 7, pp. 1361–1368, 2024, [Online]. Available: http://journal.universitaspahlawan.ac.id/index.php/jrpp/article/view/25176%0Ahttp://journal.universitaspahlawan.ac.id/index.php/jrpp/article/download/25176/17529

[9] L. J. Su and B. A. Sekti, "Implementasi Artificial Intelligence dalam Meningkatkan Cyber Security : Analisis ancaman dan Pencegahan," pp. 199–203.

[10] G. D. Setyawan, A. Yuswanto, A. M. Ridwan, B. Wibowo, and M. Firmansyah, "Implementasi Metode Adasyn Dalam Deteksi Url Berbahaya Menggunakan Machine Learning: Demi Meningkatkan Keamanan Siber Di Era Digital," *Teknokom*, vol. 6, no. 2, pp. 123–126, 2023, doi: 10.31943/teknokom.v6i2.153.

[11] J. Pebralia, "Analisis Curah Hujan Menggunakan Machine Learning Metode Regresi Linier Berganda Berbasis Python dan Jupyter Notebook," *J. Ilmu Fis. dan Pembelajarannya*, vol. 6, no. 2, pp. 23–30, 2022, doi: 10.19109/jifp.v6i2.13958.