



# Algoritma LightGBM untuk Deteksi Aktivitas *Cyber Espionage* Melalui Dataset Serangan Siber

Tasya Wulandari<sup>1\*</sup>, Yasyfi Farhan Yudisthira<sup>2</sup>, Kayla Chika P.H<sup>3</sup>, Muhammad Afriza Wibowo<sup>4</sup>, Christofer Andrew<sup>5</sup>

<sup>1,2,3,4,5</sup> Prodi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Indonesia

## Info Artikel

### Riwayat Artikel:

Diterima : **19 November 2025**

Direvisi : **18 Desember 2025**

Diterbitkan : **27 Desember 2025**

### Kata Kunci:

Espionage Siber,  
LightGBM,  
Machine Learning,  
Deteksi Awal.

## ABSTRAK

Spionase siber adalah jenis serangan siber di mana peretas mencoba mencuri informasi penting secara rahasia dan terus-menerus melalui jaringan komputer. Studi ini mengusulkan penggunaan algoritma Light Gradient Boosting Machine (LightGBM) untuk mendeteksi tanda-tanda awal spionase digital. Data yang digunakan berasal dari dataset Cybersecurity Intrusion Detection yang tersedia di Kaggle. Penelitian ini mencakup langkah-langkah seperti membersihkan dan mengorganisir data, membaginya menjadi 80% untuk pelatihan dan 20% untuk pengujian, serta melatih model dengan pengaturan yang dipilih secara hati-hati untuk laju pembelajaran dan jumlah daun. Hasil menunjukkan bahwa model LightGBM berkinerja baik, dengan akurasi 89%, AUC 0.874, dan presisi rata-rata 0.9064. Untuk kelas serangan, model memiliki presisi 1,00 dan recall 0,75. Fitur-fitur paling penting yang membantu mengidentifikasi perilaku mencurigakan adalah *ip\_reputation\_score*, *session\_duration*, dan *network\_packet\_size*. Deteksi dini dilakukan dengan menganalisis pola-pola tidak biasa dalam data untuk menemukan aktivitas jaringan yang mirip dengan spionase siber. Dibandingkan dengan metode lain seperti SVM dan Random Forest, LightGBM bekerja lebih baik dan lebih cepat. Berdasarkan hasil ini, model LightGBM dianggap sebagai alat yang baik untuk sistem peringatan dini dalam mendeteksi spionase siber.

### Keywords:

*Cyber Espionage,*  
*LightGBM,*  
*Machine Learning,*  
*Intrusion Detection.*

## ABSTRACT

*Cyber espionage is a type of cyberattack where hackers try to secretly and continuously steal important information through computer networks. This study suggests using the Light Gradient Boosting Machine (LightGBM) algorithm to spot early signs of digital espionage. The data used comes from the Cybersecurity Intrusion Detection dataset available on Kaggle. The research includes steps like cleaning and organizing the data, dividing it into 80% for training and 20% for testing, and training the model with carefully chosen settings for learning rate and number of leaves. The results show that the LightGBM model performed well, with an accuracy of 89%, an AUC of 0.874, and an average precision of 0.9064. For the attack class, the model had a precision of 1.00 and a recall of 0.75. The most important features that helped identify suspicious behavior were *ip\_reputation\_score*, *session\_duration*, and *network\_packet\_size*. Early detection happens by looking at unusual patterns in the data to find network activities that look like cyber espionage. When compared to other methods like SVM and Random Forest, LightGBM works better and is faster. Based on these results, the LightGBM model is seen as a good tool for an early warning system to detect cyber espionage.*

Tasya Wulandari,  
Program Studi Sistem Informasi,  
Universitas Bina Sarana Informatika  
Email: [19231378@bsi.ac.id](mailto:19231378@bsi.ac.id)

This is an open access article under the [CC BY-SA](#) license



## 1. PENDAHULUAN

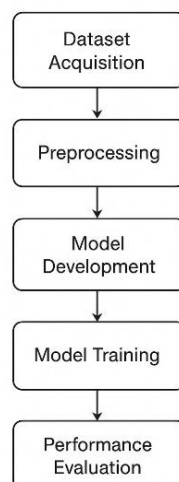
Spionase siber adalah tindakan ilegal di mana peretas mencuri informasi rahasia menggunakan jaringan komputer[1]. Di antara semua ancaman siber, spionase siber merupakan salah satu ancaman terbesar karena sulit dideteksi dan berlangsung dalam jangka waktu yang lama. Peretas menggunakan metode seperti *phishing*, memasukkan perangkat lunak berbahaya ke dalam sistem, dan Ancaman Persisten Lanjutan (APT)[2]. Menurut *Internet Crime Complaint Center (IC3)*, pada tahun 2020, Amerika Serikat mencatat lebih dari 791.000 kasus kejahatan siber, yang mengakibatkan kerugian lebih dari USD 4,2 miliar[3]. Penelitian ini fokus pada spionase siber karena meningkatnya risiko spionase digital dan kebutuhan akan metode baru untuk mendeteksi serangan baru menggunakan *machine learning*. Dampak spionase siber sangat luas. Termasuk kerugian finansial, penurunan kepercayaan terhadap pemerintah, kebocoran data penting, dan risiko terhadap keamanan dan ketertiban politik suatu negara[2]. Studi sebelumnya sebagian besar fokus pada teknik AI untuk mendeteksi ancaman siber umum[4]. Namun, sedikit penelitian tentang mendeteksi dan menghentikan spionase siber, terutama menggunakan *machine learning* untuk mengidentifikasi pola serangan tersembunyi dan aktivitas spionase digital. Studi juga menunjukkan bahwa sebagian besar dataset yang ada dirancang untuk mendeteksi serangan jaringan biasa, bukan untuk mengidentifikasi tindakan spionase spesifik[4].

Karena batasan-batasan ini, penelitian ini mengusulkan penggunaan algoritma *Light Gradient Boosting Machine (LightGBM)* untuk mendeteksi spionase siber secara dini dan efisien. *LightGBM* adalah jenis algoritma *boosting* yang membantu mempercepat proses pelatihan dan menggunakan lebih sedikit memori tanpa mengorbankan akurasi[5].

Banyak penelitian menunjukkan bahwa *LightGBM* berkinerja lebih baik daripada algoritma lain seperti *XGBoost* dan *Random Forest* saat bekerja dengan data keamanan siber yang besar[6]. Algoritma ini mampu menangani fitur yang kompleks dan data yang tidak seimbang, menjadikannya sangat baik untuk mendeteksi tanda-tanda awal spionase digital, yang memerlukan analisis cepat dan akurat terhadap pola serangan tersembunyi. Di masa lalu, metode seperti *Support Vector Machine (SVM)*, *Random Forest*, dan *Logistic Regression* telah digunakan dalam Sistem Deteksi Intrusi (IDS). Namun, metode-metode ini masih kesulitan dalam mendeteksi serangan non-linear, mengikuti kecepatan pemrosesan yang tinggi, dan skalabilitas yang tepat saat menangani dataset besar[7]. Teknik klasifikasi tradisional juga kesulitan mengikuti sifat canggih dan kompleks serangan spionase siber modern.

Bagian baru dari studi ini adalah penggunaan *LightGBM* untuk deteksi dini spionase digital, metode yang belum banyak diteliti dalam keamanan siber. Penelitian ini menyoroti kebutuhan akan dataset yang lebih terfokus dengan beberapa fitur yang relevan untuk mendeteksi spionase dan model yang dirancang khusus untuk identifikasi dini. Pendekatan ini diharapkan dapat meningkatkan sistem peringatan dini terhadap spionase siber dan mendukung pengembangan sistem keamanan. Dengan memahami ancaman ini secara lebih baik, organisasi dapat mengambil langkah-langkah yang lebih kuat untuk mencegahnya[8].

## 2. METODE PENELITIAN



Gambar 1. Alur Penelitian

Penelitian ini terdiri dari beberapa tahap, mulai dari identifikasi dataset, prapemrosesan, pengembangan model, pelatihan, hingga evaluasi kinerja. Alur keseluruhan dari metodologi yang diusulkan dirancang untuk mendeteksi aktivitas spionase siber berdasarkan pola perilaku lalu lintas jaringan menggunakan algoritma *LightGBM* seperti yang ditunjukkan pada Gambar 1.

Data set yang digunakan dalam studi ini adalah *Cybersecurity Intrusion Detection Dataset* yang diperoleh dari Kaggle, berisi 9.537 catatan dan 9 atribut fitur yang mewakili aktivitas jaringan normal dan berbahaya. Atribut-atribut ini meliputi *network\_packet\_size*, *protocol\_type*, *login\_attempts*, *session\_duration*, *encryption\_used*, *ip\_reputation\_score*, *failed\_logins*, *browser\_type*, dan *unusual\_time\_access*, yang menjadi dasar untuk mengidentifikasi pola spionase siber yang mencurigakan. Sebelum memulai pelatihan model, data dipersiapkan untuk memastikan konsistensi dan kesiapannya untuk digunakan dengan algoritma pembelajaran mesin. Proses prapemrosesan, yang merupakan proses mengubah data tidak terstruktur atau mentah menjadi format terstruktur yang sesuai untuk pelatihan model[9], dimulai dengan pembersihan data. Langkah ini berfokus pada penanganan nilai yang hilang pada atribut penting untuk menjaga kualitas dataset dan memastikan semua catatan Tetap dapat digunakan[10]. Hal ini melibatkan penghapusan nilai yang hilang dan kolom yang tidak perlu seperti *session\_id* karena kolom-kolom tersebut tidak berkontribusi pada prediksi dan hanya berfungsi sebagai penanda identifikasi. Kemudian, data diubah menggunakan metode yang disebut *Label Encoding* untuk mengubah kategori seperti jenis protokol atau jenis browser menjadi angka. *Label Encoding* diterapkan menggunakan metode *LabelEncoder*[11], [12], yang mengubah setiap kategori teks unik menjadi nilai numerik yang berbeda, yang dapat digunakan oleh algoritma *LightGBM*. Setelah itu, data dibagi menjadi dua bagian: fitur (X) dan hasil target (y). Data kemudian dibagi menjadi set pelatihan dan pengujian dengan rasio 80:20 menggunakan fungsi *train\_test\_split*[13], dan parameter *stratify* digunakan untuk memastikan distribusi kelas tetap seimbang di kedua set.

*Light Gradient Boosting Machine (LightGBM)* dipilih karena performanya yang baik pada dataset besar dengan banyak fitur dan distribusi yang tidak merata. Algoritma ini banyak digunakan untuk tugas klasifikasi berskala besar berkat kecepatan, kemampuannya menangani data berdimensi tinggi, dan penggunaan memori yang efisien[14]. *LightGBM* menggunakan metode *boosting* berbasis gradien dan meningkatkan kecepatan pelatihan dengan menggunakan metode histogram tanpa mengorbankan akurasi. Model ini dilatih dengan pengaturan spesifik, seperti laju pembelajaran 0.05, 31 *Num\_leaves* di setiap pohon, fraksi fitur 0.9, dan 100 putaran *boosting*. Studi sebelumnya juga menyoroti pentingnya menyesuaikan *num\_leaves*[15] dan laju pembelajaran[16] untuk meningkatkan kinerja deteksi, terutama pada dataset jaringan yang tidak seimbang. Pengaturan ini membantu model mendeteksi pola serangan yang kompleks yang sering ditemukan dalam aktivitas spionase siber.

Setelah pelatihan model, evaluasi kinerja dilakukan untuk menilai kemampuan model dalam mengklasifikasikan perilaku spionase siber. Beberapa metrik evaluasi digunakan, termasuk akurasi, presisi, *recall*, dan *F1-score*, yang merupakan rata-rata harmonik dari presisi dan *recall*, memberikan evaluasi yang seimbang[17]. Metrik-metrik ini memberikan pemahaman komprehensif tentang kinerja model. Matriks kebingungan juga digunakan untuk menganalisis kesalahan klasifikasi dengan membandingkan label yang diprediksi dengan label sebenarnya, digunakan untuk memvisualisasikan hasil klasifikasi dengan menunjukkan jumlah *true positives (TP)*, *true negatives (TN)*, *false positives (FP)*, dan *false negatives (FN)*[18] seperti yang ditunjukkan pada Gambar 2.

## Confusion Matrix

|                        | Actually Positive (1) | Actually Negative (0) |
|------------------------|-----------------------|-----------------------|
| Predicted Positive (1) | True Positives (TPs)  | False Positives (FPs) |
| Predicted Negative (0) | False Negatives (FNs) | True Negatives (TNs)  |

Gambar 2. Confusion Matrix

Evaluasi tambahan dilakukan menggunakan kurva ROC-AUC dan *Precision-Recall*, yang memberikan wawasan tentang kemampuan diskriminatif dan stabilitas model pada dataset deteksi intrusi yang tidak seimbang. Kombinasi tahap-tahap ini memungkinkan model LightGBM untuk secara efektif mengenali pola perilaku yang terkait dengan aktivitas spionase siber, yang berfungsi sebagai landasan untuk mengembangkan sistem deteksi peringatan dini dalam lingkungan keamanan siber.

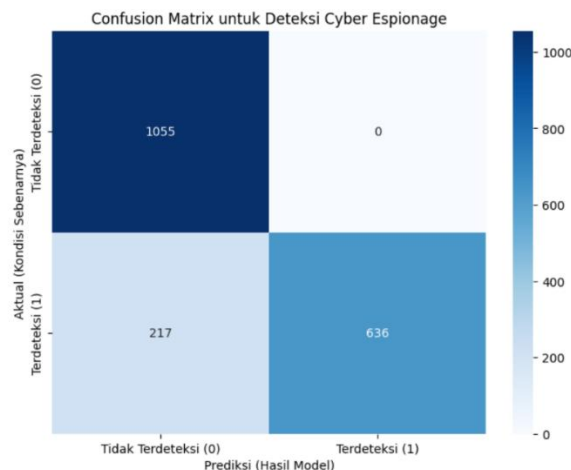
### 3. HASIL DAN ANALISIS

Model *LightGBM* dilatih menggunakan dataset yang telah diproses dan dievaluasi menggunakan beberapa metrik kinerja, termasuk akurasi, presisi, *recall*, *F1-score*, *AUC*, dan *Average Precision*. Evaluasi secara keseluruhan menunjukkan bahwa model tersebut berfungsi secara efektif dalam mendeteksi aktivitas spionase siber berdasarkan pola perilaku jaringan. Model tersebut mencapai akurasi sebesar 89%, seperti yang tercantum dalam Tabel 1, menunjukkan bahwa sebagian besar prediksi diklasifikasikan dengan benar.

**Tabel 1. Hasil klasifikasi *LightGBM***

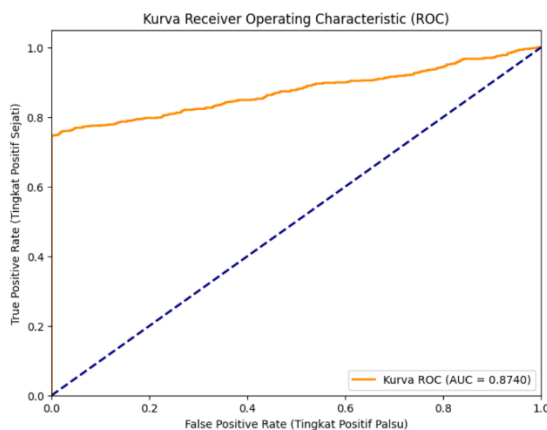
| class        | Precision | Recall | F1-Score | Support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.83      | 1.00   | 0.91     | 1055    |
| 1            | 1.00      | 0,75   | 0.85     | 853     |
| Accuracy     |           |        | 0.89     | 1908    |
| Macro Avg    | 0.91      | 0.87   | 0.88     | 1908    |
| Weighted Avg | 0.91      | 0.89   | 0.88     | 1908    |

Analisis mendalam tentang cara model menangani kasus serangan menunjukkan akurasi sebesar 1.00, artinya setiap sesi yang dilabeli sebagai serangan memang benar-benar serangan yang nyata. Hal ini sangat penting dalam keamanan siber karena alarm palsu dapat menyebabkan tim membuang waktu untuk menanggapi masalah yang sebenarnya tidak ada. Namun, nilai *recall* model sebesar 0.75 menunjukkan bahwa beberapa serangan sebenarnya terlewatkan. Hal ini menunjukkan bahwa model cukup baik dalam mendeteksi ancaman nyata, tetapi masih perlu ditingkatkan untuk mendeteksi bentuk-bentuk spionase yang lebih tersembunyi atau halus. Selain itu, Matriks Kebingungan pada Gambar 3 di bawah ini menunjukkan bahwa model berhasil mengklasifikasikan 1.691 dari 1.908 data uji dengan benar. Tidak ada kesalahan tipe I (*false positives* = 0), tetapi ada 217 kesalahan tipe II (*false negatives*), yang merupakan serangan yang tidak terdeteksi.

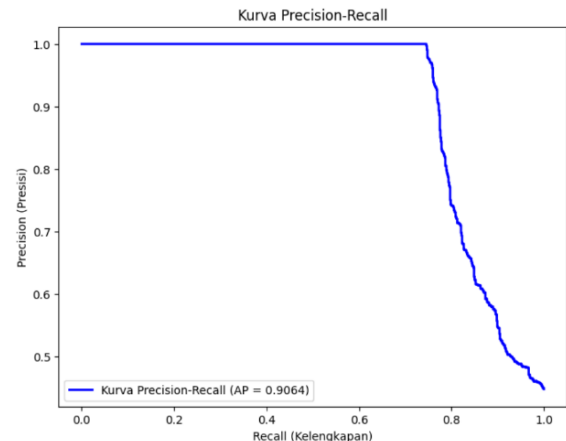


**Gambar 3. Hasil Confusion Matrix**

Model ini juga menghasilkan skor AUC sebesar 0.874, seperti yang ditunjukkan pada Gambar 4, yang menunjukkan kemampuan diskriminatif yang kuat dalam membedakan antara aktivitas yang tidak berbahaya dan berbahaya. Selain itu, skor *Average Precision* sebesar 0.9064 pada Gambar 5 menyoroti ketahanan model terhadap data yang tidak seimbang.



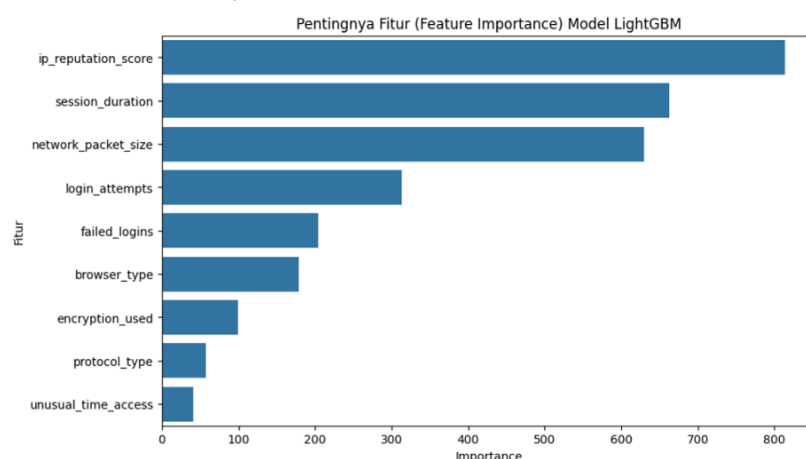
Gambar 4. Kurva AUC



Gambar 5. Kurva Average Precision

AP sangat penting dalam dataset keamanan jaringan yang tidak seimbang. Nilai AP yang tinggi menunjukkan bahwa meskipun tingkat *recall* tidak sempurna, model masih mampu mendeteksi serangan dengan presisi yang baik[19]. Hal ini merupakan karakteristik umum dari dataset deteksi intrusi. Hasil model *LightGBM* menunjukkan bahwa alur kerja prapemrosesan secara signifikan meningkatkan tingkat *recall* dibandingkan dengan dataset asli. Nilai *recall* ini lebih rendah dibandingkan dengan hasil yang diperoleh dari beberapa studi *LightGBM* pada dataset dan alur kerja prapemrosesan[20]. Situasi ini sering terjadi ketika pola serangan terlalu mirip dengan aktivitas jaringan normal, sehingga model kesulitan membedakan keduanya. Serangan sering memiliki pola yang halus dan sering menyerupai aktivitas jaringan normal untuk menghindari deteksi[21].

Analisis pentingnya fitur menunjukkan bahwa beberapa atribut secara signifikan mempengaruhi proses pengambilan keputusan model. Atribut yang paling berpengaruh adalah *ip\_reputation\_score*, *session\_duration*, dan *network\_packet\_size*, seperti yang ditunjukkan pada Gambar 6. Atribut-atribut ini memberikan indikator kuat tentang perilaku abnormal yang sering dikaitkan dengan spionase siber, seperti sesi rahasia yang lama, ukuran paket yang mencurigakan, atau akses dari *host* dengan skor reputasi yang buruk. Temuan ini konsisten dengan studi sebelumnya yang menekankan indikator berbasis perilaku sebagai elemen kritis dalam mengidentifikasi aktivitas spionase.



Gambar 6. Fitur Importance

Perbandingan menunjukkan bahwa *LightGBM* outperformed *SVM* dan model *Random Forest* dalam hal akurasi, kecepatan, dan penanganan dataset besar. Dibandingkan dengan studi sebelumnya yang menggunakan metode *SVM* dan *Random Forest*[16] dengan akurasi rata-rata sekitar 80–85%, model *LightGBM* dalam studi ini menunjukkan peningkatan kinerja yang signifikan. *SVM* lebih lambat dalam pemrosesan, dan *Random Forest* mengalami kesulitan mempertahankan akurasi saat menangani data yang tidak seimbang. Penggunaan histogram dan pendekatan pertumbuhan *leaf-wise LightGBM* membantu model ini mendeteksi pola tersembunyi yang terkait dengan serangan spionase secara lebih efektif. Secara umum, temuan ini

menunjukkan bahwa *LightGBM* adalah metode yang baik dan cepat untuk deteksi dini spionase siber. Kemampuannya untuk menemukan tanda-tanda kecil aktivitas yang tidak biasa dan mempertahankan akurasi tinggi menjadikannya pilihan yang kuat untuk digunakan dalam sistem peringatan dini. Hal ini memungkinkan organisasi untuk mendeteksi upaya spionase potensial sebelum menyebabkan masalah keamanan yang serius.

#### 4. KESIMPULAN

Studi ini menggunakan algoritma *LightGBM* untuk mendeteksi spionase siber dengan menganalisis dataset *Cybersecurity Intrusion Detection*. Setelah membersihkan dan memodifikasi data, serta melatih model, *LightGBM* mampu mempelajari pola perilaku jaringan yang normal dan mengidentifikasi tanda-tanda serangan spionase berdasarkan faktor-faktor seperti skor reputasi IP, durasi sesi, dan ukuran paket jaringan. Hasil menunjukkan bahwa *LightGBM* memiliki akurasi 89%, presisi sempurna untuk kategori serangan, dan skor AUC 0.874, yang berarti model dapat membedakan antara aktivitas normal dan berbahaya dengan cukup baik. Namun, skor recall hanya 0,75, artinya beberapa serangan tidak terdeteksi. Hal ini menunjukkan adanya false negative, kemungkinan karena lalu lintas normal dan aktivitas spionase dapat terlihat serupa. Ini menunjukkan bahwa meskipun *LightGBM* bekerja baik dengan pola serangan yang diketahui, model ini mungkin melewatkan aktivitas spionase yang lebih tersembunyi atau diam-diam.

Secara umum, hasil menunjukkan bahwa *LightGBM* adalah alat yang baik untuk deteksi dini spionase siber dan dapat menjadi bagian dari sistem keamanan jaringan yang memberikan peringatan dini. Perlu dilakukan penelitian lebih lanjut untuk meningkatkan kemampuan model dalam mendeteksi serangan yang halus, mengurangi false negatives, dan menganalisis dataset yang lebih beragam atau spesifik untuk membuat deteksi lebih akurat dan andal.

#### REFERENSI

- [1] A. Ciluvi, S. Luma-Osmani, E. Rufati, and G. Arifi, "Cyber Espionage: A Growing Threat?," *J. Nat. Sci. Math.*, vol. 7, no. 13/14, pp. 12–20, 2022, [Online]. Available: [http://nuuance.net/work/Cyber Espionage.pdf](http://nuuance.net/work/Cyber%20Espionage.pdf)
- [2] M. E. Susila and A. A. Salim, "Cyber Espionage Policy and Regulation: A Comparative Analysis of Indonesia and Germany A . Introduction As technology evolves , it supersedes traditional forms of communication and information technology with newer , often more efficient , media . 1 Amon," *Pjih*, vol. 11, no. 1, pp. 9–11, 2024.
- [3] "2020 Internet Crime Report," pp. 1–30.
- [4] E. Y. Fitria, K. Mutijarsa, P. Korespondensi, S. Siber, K. Buatan, and D. Learning, "RESEARCH SURVEY ON ARTIFICIAL INTELLIGENCE METHODS FOR DETECTING CYBERATTACK TECHNOLOGY THREATS," vol. 10, no. 6, 2023, doi: 10.25126/jtiik.2023107341.
- [5] M. Atta, O. Ahmed, Y. Abdelsatar, R. Alotaibi, and O. Reyad, "Enhancing Internet of Things security using performance gradient boosting for network intrusion detection systems," *Alexandria Eng. J.*, vol. 116, no. December 2024, pp. 472–482, 2025, doi: 10.1016/j.aej.2024.12.106.
- [6] M. Nidhi, S. Kadam, and M. Gayakwad, "An Ensemble Model for Cyber Attack and Threat Detection in Applications Network Using Random Forest , Lightgbm and Xgboost," vol. 28, no. 3, pp. 523–534, 2025.
- [7] E. Ramadanti, D. Aprilya Dinathi, C. Christianskaditya, and D. R. Chandranegara, "Diabetes Disease Detection Classification Using Light Gradient Boosting (LightGBM) With Hyperparameter Tuning," *Sinkron*, vol. 8, no. 2, pp. 956–963, 2024, doi: 10.33395/sinkron.v8i2.13530.
- [8] T. G. Laksana, S. Mulyani, I. Komputer, J. Informatika, and U. Bhayangkara, "UNTUK MENCEGAH PEMBOBOLAN DATA PERUSAHAAN Contact: Cite This Article:," vol. 3, no. 1, pp. 109–122, 2024.
- [9] J. N. Sibarani, D. R. Sirait, and S. Ramadhanti, "Intrusion Detection Systems pada Bot-IoT Dataset Menggunakan Algoritma Machine Learning," vol. 14, no. 1, pp. 38–52, 2023.
- [10] J. Nasional and S. Informasi, "Integrasi Algoritma FP-Growth dan K-Means untuk Analisis Keranjang Belanja dan Segmentasi Pelanggan pada Data Transaksi Ritel," vol. 02, pp. 128–135, 2025.
- [11] W. Yustanti and N. Iriawan, "Categorical encoder based performance comparison in pre- processing

- imbalanced multiclass classification," vol. 31, no. 3, pp. 1705–1715, 2023, doi: 10.11591/ijeecs.v31.i3.pp1705-1715.
- [12] W. Zhu, R. Qiu, and Y. Fu, "Comparative Study on the Performance of Categorical Variable Encoders in Classification and Regression Tasks," pp. 1–19.
- [13] J. Nasional, S. Informasi, E. Mustika, C. Sabila, R. Fakhrizal, and R. Kurniawan, "Analisis dan Prediksi Indeks Kualitas Udara Jakarta: Penerapan Algoritma XGBoost," vol. 02, pp. 161–169, 2025.
- [14] F. B. Dergisi, "A new Intrusion Detection System for Secured IoT / IIoT Networks based on A new Intrusion Detection System for Secured IoT / IIoT Networks based on Güvenli IoT / IIoT Ağları İç in LGBM T abanlı Yeni B ir Saldırı Tespit Sistemi," 2023, doi: 10.29109/gujsc.1173286.
- [15] Y. Cong, T. Guan, J. Cui, and X. Cheng, "LGBM: An Intrusion Detection Scheme for Resource-Constrained End Devices in Internet of Things," vol. 2022, no. 1, 2022, doi: 10.1155/2022/1761655.
- [16] G. Ketepalli, S. Y. M, and K. N. S. Lakshmi, "A Hybrid Intrusion Detection Model Using LSTMAE and LightGBM for Robust Anomaly Detection in Network Systems," vol. 14, no. 5, pp. 912–922, 2025.
- [17] R. Yacoub, "Probabilistic Extension of Precision , Recall , and F1 Score for More Thorough Evaluation of Classification Models," pp. 79–91, 2020.
- [18] J. Crall, "AND RECALL AS TRUE NEGATIVES APPROACH INFINITY .," no. 2, 2023.
- [19] N. Tran, H. Chen, J. Jiang, J. Bhuyan, and J. Ding, "Effect of Class Imbalance on the Performance of Machine Learning-based Network Intrusion Detection," vol. 17, no. 9, pp. 741–755, 2021, doi: 10.23940/ijpe.21.09.p1.741755.
- [20] A. Technology, "UNITEC INSTITUTE OF TECHNOLOGY MASTER THESIS A Comparative Analysis of Machine Learning Algorithms in Network-Based Intrusion Detection Systems for Detecting Advanced Persistent Threats to Enhance Cybersecurity A thesis submitted in partial fulfilment of the requirements for the Degree of Master of Applied Technologies in the School of Computing , Electrical and Applied Technology," 2024.
- [21] T. Ngootip, "Enhancing Network Intrusion Detection in Cloud Computing Using a Deep Boltzmann Machine and LightGBM Ensemble Model: A Performance Evaluation on the NSL-KDD Dataset," vol. 1, no. 1, pp. 1–7, 2024.