Vol. 6, No. 3, Desember 2025, hal. 568-574

Jurnal Computer Science and Information Technology (CoSciTech)

p-ISSN: 2723-567X
e-ISSN: 2723-5661

http://ejurnal.umri.ac.id/index.php/coscitech/index

Deteksi Serangan Dalam Ekosistem Iot Melalui Analisis Multi-Class Dengan Model Xgboost Dan Penerapan Teknik Imbalance Ratio Pada Dataset IoTID20

Januar Al Amien*[1], Sunanto[2], Muhammad Al-Ikhsan Rangkuti[3], Soni[4]
Email: [1]januaralamien@umri.ac.id, [2]sunanto@umri.ac.id, [3]200401071@student.umri.ac.id,[4]soni@umri.ac.id

[1234]Teknik Informatika, Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau

Diterima: 29 Juli 2025 | Direvisi: 06 September 2025 | Disetujui: 25 Desember 2025
©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Penelitian ini berfokus pada deteksi serangan dalam ekosistem *Internet of Things* (IoT) menggunakan algoritma *XGBoost* dan teknik *Imbalance Ratio* pada dataset IoTID20. Tujuan utama adalah mengatasi masalah ketidakseimbangan data yang umum terjadi pada dataset *Intrusion Detection System* (IDS) dan meningkatkan akurasi dalam mengklasifikasikan jenis serangan. Metodologi yang digunakan meliputi preprocessing data, seleksi fitur, dan penerapan teknik *Imbalance Ratio* untuk menangani ketidakseimbangan kelas pada dataset IoTID20. Selanjutnya, model *XGBoost* diimplementasikan dengan parameter *scale_pos_weight* untuk menangani masalah ketidakseimbangan kelas. Model ini dilatih pada data training dan dievaluasi menggunakan metrik seperti akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa kombinasi algoritma *XGBoost* dan teknik *Imbalance Ratio* mampu mengatasi masalah ketidakseimbangan data secara efektif. Model yang dihasilkan mencapai tingkat akurasi 99,32%, presisi 99,32%, recall 99,32%, dan F1-score 99,32% dalam mengklasifikasikan jenis serangan pada dataset IoTID20. Hasil ini menunjukkan kemampuan yang sangat baik dalam mendeteksi serangan dan membedakan antara lalu lintas normal dan anomali dalam ekosistem *IoT*. Penelitian ini memberikan kontribusi dalam meningkatkan keamanan jaringan *IoT* dengan menerapkan pendekatan *Machine Learning* yang efektif untuk mendeteksi serangan secara akurat, sekaligus menangani masalah ketidakseimbangan data yang sering terjadi pada dataset *Intrusion Detection System* (IDS)

Kata kunci: *Internet of Things (IoT), XGBoost, Imbalance Ratio, Ketidakseimbangan Data, Deteksi Serangan, Dataset IoTID20, Klasifikasi*.

Attack Detection in the IoT Ecosystem Through Multi-Class Analysis with the Xgboost Model and Application of the Imbalance Ratio Technique on the IoTID20 Dataset

Abstract

This research focuses on attack detection in the Internet of Things (IoT) ecosystem using the XGBoost algorithm and the Imbalance Ratio technique on the IoTID20 dataset. The main goal is to overcome the problem of data imbalance that is common in IDS datasets and improve accuracy in classifying attack types. The methodology used includes data preprocessing, feature selection, and applying the Imbalance Ratio technique to handle class imbalance in the IoTID20 dataset. Next, the XGBoost model is implemented with the scale_pos_weight parameter to handle the class imbalance problem. This model is trained on training data and evaluated using metrics such as accuracy, precision, recall, and F1-score. The research results show that the combination of the XGBoost algorithm and the Imbalance Ratio technique is able to overcome data imbalance problems effectively. The resulting model achieved an accuracy rate of 99.32%, precision 99.32%, recall 99.32%, and F1-score 99.32% in classifying attack types on the IoTID20 dataset. These results demonstrate excellent capabilities in detecting attacks and distinguishing between normal and anomalous traffic in the IoT ecosystem. This research contributes to improving IoT network security by applying an effective Machine Learning approach to accurately detect attacks, while also addressing data imbalance problems that often occur in IDS datasets.

Keywords: *Internet of Things (IoT), XGBoost, Imbalance Ratio, Data Imbalance, Attack Detection, IoTID20 Dataset, Classification*

1. PENDAHULUAN

Dengan pesatnya kemajuan teknologi informasi, penggunaan Internet of Things (IoT) telah merambah ke segala bidang. Menurut (Khraisat et al. 2019), Internet of Things (IoT) merupakan sistem perangkat yang saling terkoneksi dan saling bertukar informasi melalui jaringan internet. Belakangan ini, Internet of Things (IoT) telah menarik perhatian dalam komunitas penelitan dan industri karena manfaatnya [1].

Namun, memastikan keamanan dan privasi perangkat IoT merupakan tantangan yang tidak bisa dianggap remeh, karena keterbatasan kemampuan komputasi perangkat IoT yang tidak cukup untuk mekanisme keamanan tradisional. Hal ini membuat perangkat IoT rentan terhadap serangan seperti kebocoran data, spoofing, dan DoS/DDoS [2]. Untuk mengatasi hal keamanan tersebut, digunakan sebuah teknologi bernama Intrusion Detection System (IDS). IDS terus-menerus memantau lalu lintas jaringan yang berasal dari berbagai sumber untuk mendeteksi kelainan, yang mungkin merupakan ancaman keamanan [3].

Intrusion Detection System (IDS) adalah sebuah sistem yang melakukan pemantauan berbagai sistem jaringan [4]. IDS mampu mengawasi serta mendeteksi serangan yang mengancam fitur keamanan (kerahasiaan, ketersediaan dan integritas) suatu sistem [5]. Informasi seperti lalu lintas jaringan log keamanan dan serangan, tercakup didalam IDS yang dirangkum dalam bentuk dataset [6]. Terdapat banyak peneliti yang menggunakan dataset-dataset IDS untuk melakukan pengembangan IDS dengan pendekatan Machine Learning. Seperti pada penelitian [7] menggunakan dataset KDD-CUP99 menggunakan algoritma CNN dan LSTM. Kemudian pada penelitian [8] menggunakan dataset NSL-KDD menggunakan algoritma KNN, Random Forest, Gradient Boosted Tree dan Decision Tree. Penelitian berikutnya [9] dataset UNSW-NB15 dengan menggunakan algoritma XGBoost. [10] menggunakan dataset BoT-IoT dengan pendekatan algoritma KNN. [11] menggunakan dataset TON_IoT dengan pendekatan algoritma XGBoost. pada penelitian [12] menggunakan dataset BotIoT dengan pendekatan algoritma XGBoost. Pada penelitian [13] memperkenalkan dataset baru yang diberi nama IoTID20, kemudian dilakukan pendekatan Machine Learning algoritma Ensemble. Dalam penelitian ini penulis merujuk pada dataset IoTID20 dengan pendekatan algoritma XGBoost.

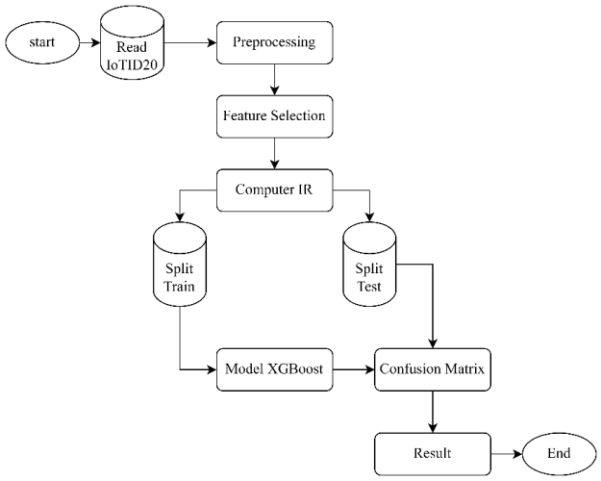
Dari sekian banyak penelitian diatas, terdapat masalah imbalance data pada dataset. Banyak dataset IDS yang mempunyai masalah ketidakseimbangan data/ imbalance data yang mengakibatkan penurunan tingkat akurasi pada algoritma termasuk dataset IoTID20 yang digunakan pada penelitian ini [14]. Pada dataset yang tidak seimbang, kelas yang memiliki jumlah lebih banyak disebut kelas mayoritas, sedangkan kelas yang memiliki jumlah yang relatif lebih sedikit disebut kelas minoritas [15]. Akibat dari data yang tidak seimbang tersebut, algoritma tidak mempelajari data kelas minoritas dengan baik dan cenderung bias terhadap kelas mayoritas yang merupakan sekumpulan data besar [16].

Untuk mengatasi masalah ketidakseimbangan kelas tersebut, digunakan cara untuk menyeimbangkan data yang tidak seimbang dengan menyesuaikan jumlah data di salah satu kelas. Teknik tersebut Bernama Undersampling dan Oversampling [17]. Pada penelitian [18] menggunakan teknik Oversampling Adaptive Synthetic Sampling (ADASYN) dengan menaikkan jumlah sampel kelas minoritasnya. [19] menggunakan teknik Synthetic Minority Oversampling Technique (SMOTE) untuk mengatasi ketidakseimbangan kelas pada dataset. Kemudian pada penelitian [20], menggunakan teknik Radial-Based Undersampling (RBU) dengan algoritma Decision Tree. [12] menggunakan teknik Imbalance Ratio untuk mengatasi ketidakseimbangan kelas pada dateset BotIoT.

Penelitian ini berfokus pada strategi penanganan ketidakseimbangan data dalam dataset IoTID20 dengan memanfaatkan metode Imbalance Ratio untuk menentukan bobot yang optimal bagi setiap kelas, sehingga memungkinkan model memberikan prioritas pada kelas minor. Tujuannya adalah untuk meningkatkan akurasi deteksi berbagai jenis serangan melalui implementasi algoritma Machine Learning XGBoost. Dengan demikian, penelitian ini bertujuan untuk mengoptimalkan kinerja model dalam mengidentifikasi serangan pada jaringan IoT dengan pendekatan yang fokus pada penyelesaian masalah ketidakseimbangan kelas.

2. METODE PENELITIAN

Adapun metode penelitian atau langkah-langkah yang dilakukan dalam penelitian ini, mencakup alur kerja penelitian, proses pengumpulan data, analisis dataset, pengembangan dataset menggunakan pendekatan machine learning, penerapan metode XGBoost, tahap implementasi dan pengujian. Selain itu, juga membahas tentang pemahaman terhadap sistem pemikiran yang kompleks dengan pendekatan pemecahan berdasarkan unsur-unsur penelitian tersebut, sehingga membentuk suatu alur kerja yang terstruktur.



Gambar 1. Metodologi Penelitian

2.1. Dataset IoTID20

Dataset IoTID20 mencakup berbagai jenis serangan pada Internet of Things (IoT), termasuk DDoS, DoS, Mirai, ARP Spoofing, dan lain sebagainya, bersama dengan lalu lintas normal. Kumpulan data ini dikumpulkan dari ekosistem IoT di sebuah rumah pintar yang dirancang untuk mengintegrasikan berbagai komponen terhubung, termasuk Speaker AI (SKTNGU), kamera Wi-Fi (EZVIZ), laptop, ponsel pintar, tablet, titik akses nirkabel (Wi-Fi), dan router Wi-Fi. Dalam konteks dataset ini, kamera dan speaker AI diwakili sebagai perangkat IoT yang menjadi korban, sementara perangkat lain diwakili sebagai perangkat penyerang. Sebuah testbed telah diimplementasikan untuk mensimulasikan berbagai serangan yang mungkin terjadi di ekosistem IoT, menggunakan alat Network Mapper (Nmap). Dataset IoTID20 berisi 625.784 *row*/data dan 86 *column*/fitur. Kolom dataset yang digunakan pada penelitian ini yaitu 'Cat' yang bertipe *object*. Pada kolom 'Cat' berisi 5 kelas yaitu *Mirai*, *Scan*, *DoS*, *Normal*, *MITM ARP Spoofing*.

Table 1. Kelas dan Serangan Dataset IoTID20

No	Tipe	Number
1	Normal	40073
2	DoS	59391
3	Mirai	415677
4	MITM	35377
5	Scan	75265

2.2. Preprocessing

Dalam proses pra-pemrosesan data, langkah pertama adalah menerapkan label encoding pada fitur kategorikal seperti Flow_ID, Src_IP, dan Dst_IP untuk mengubahnya menjadi representasi numerik yang dapat dipahami oleh model. Selanjutnya, penggunaan Robust Scaler menjadi penting untuk menangani nilai infinit yang mungkin ada dalam fitur Flow_Byts/s dan Flow_Pkts/s, menghindari potensi masalah numerik dalam pemrosesan data. Langkah terakhir adalah menghapus fitur Flow_ID dan Timestamp dari dataset, karena mereka mungkin tidak memberikan informasi yang relevan untuk proses pembelajaran mesin yang diinginkan [5]. Dengan langkah-langkah ini, data siap untuk digunakan dalam membangun model pembelajaran mesin yang akurat dan andal.

2.3. Feature Selection

Proses Feature Selection bertujuan untuk mengidentifikasi dan memisahkan fitur-fitur yang paling relevan dalam dataset. Fitur-fitur yang dianggap tidak diperlukan seperti 'Flow_ID', 'Timestamp', 'Label', dan 'Sub_Cat' dihapus dari dataset menggunakan metode drop, memperbaiki kualitas data untuk analisis lebih lanjut. Dengan demikian, hanya fitur-fitur yang memberikan kontribusi signifikan terhadap analisis yang dipertahankan dalam proses seleksi fitur [21].

2.4. Compute Imbalance Ratio

Algoritma ini digunakan untuk menghitung nilai Imbalance Ratio (IR) dari sebuah dataset yang memiliki beberapa kelas. Tujuan dari algoritma dibawah untuk menekankan atau memberikan perhatian agar model memberikan bobot besar kepada kelas minor

[22]. Pertama, jumlah sampel dalam setiap kelas dihitung. Kemudian, untuk setiap kelas, IR dihitung sebagai rasio antara jumlah sampel dalam kelas mayoritas dan kelas minoritas. Selanjutnya, FIR (maximum imbalance ratio) dihitung sebagai nilai maksimum dari semua IR kelas. Langkah selanjutnya adalah menghitung rata-rata nilai yang diperoleh pada langkah sebelumnya. Hasil akhir dari algoritma ini adalah nilai rata-rata dari IR di setiap kelas, memberikan pemahaman tentang tingkat ketidakseimbangan dataset tersebut.

Find the number of samples in each class

For each class i, calculate the number of samples in the majority class (N_i) and the number of samples in the minority class (n_i).

Calculate the imbalance ratio (IR_i) for each class i as follows:

$$IR\_i = N\_i / n\_i \quad (1)$$

Calculate the FIR value for the dataset as the maximum imbalance ratio across all classes:

$$FIR = max(IR\_1, IR\_2, \ldots, IR\_k) \quad (2)$$

Calculate the average of the values obtained in step 2

Return the result

where k is the total number of classes in the dataset.

2.6. Model XGBoost

Pembagian data dibagi menjadi training 80% dan testing 20% Dalam implementasi model *XGBoost*, *XGBClassifier* dari pustaka *XGBoost* digunakan untuk melakukan klasifikasi. Awalnya, kelas `*XGBClassifier*` dari pustaka *XGBoost* diimpor. Waktu awal eksekusi diinisialisasi menggunakan `time.time()` dan pengumpulan sampah diaktifkan menggunakan `gc.enable()`. Model *XGBoost* dibuat dengan parameter *scale_pos_weight* untuk menangani ketidakseimbangan kelas. Model dilatih pada data pelatihan dengan metode `.fit()` dan set evaluasi ditentukan untuk memantau performa model. Prediksi kemudian dibuat pada data pengujian menggunakan `.predict()`. Waktu total eksekusi direkam dan jumlah sampah yang terkumpul diatur ulang sebelum kode selesai dieksekusi.

2.7. Evaluasi (Confusion Matrix)

Dalam penelitian ini, pengujian dilakukan menggunakan model *XGBoost* dengan menggunakan dataset IoTID20. Dengan menggunakan *Confusion Matrix*, kita dapat menghitung beberapa metrik evaluasi, seperti akurasi, presisi, recall (sensitivitas), dan F1-score [23]. *Confusion Matrix* ini berguna untuk menghasilkan berbagai metrik dengan memadukan nilai *True Negative* (TN), *True Positive* (TP), *False Negative* (FN), dan *False Positive* (FP) [24]. Pemanfaatan Confusion Matrix dalam evaluasi memungkinkan penilaian terhadap kemampuan model klasifikasi dalam mengklasifikasikan data secara tepat, serta mengidentifikasi area yang memerlukan perbaikan atau penyesuaian dalam model.

3. HASIL DAN PEMBAHASAN

3.1. Imbalance Ratio

Tabel 2. menunjukkan kelas serangan dalam dataset, yang mencakup jumlah *instance* dari masing-masing kelas serta nilai *Imbalance Ratio* (IR) yang menggambarkan tingkat ketidakseimbangan antara jumlah *instance* dari kelas mayor dan kelas minor. Kelas Mirai memiliki IR terendah dengan nilai 0.301, menandakan ketidakseimbangan yang paling signifikan dalam distribusi kelas, sementara kelas Normal memiliki IR tertinggi dengan nilai 3.123, menunjukkan distribusi yang lebih seimbang antara kelas mayor dan kelas minor. Dengan demikian, analisis ini memberikan wawasan tentang distribusi relatif dari setiap kelas serangan dalam dataset.

Table 2. Imbalance Ratio

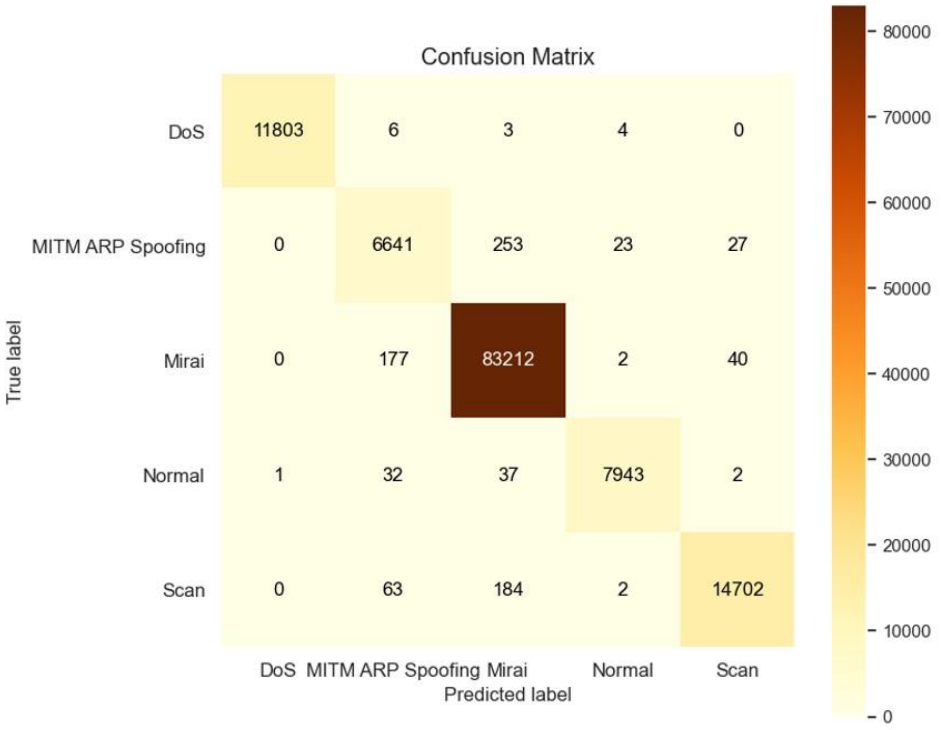
No	Nama Class	Class Of Number	Class Of Number IR
1	DoS	59391	2.107332760856022
2	MITM ARP Spoofing	35377	3.5377957429968623
3	Mirai	415677	0.30109099132258943
4	Normal	40073	3.1232151323834003
5	Scan	75265	1.6628791603002724

3.2. XGBoost

Tabel 3 menunjukkan hasil evaluasi performa model XGBoost dalam mengklasifikasikan kelima kelas serangan yang terdapat dalam dataset IoTID20 menggunakan metrik evaluasi, yaitu precision, recall, F1-score, dan overall accuracy. Precision mengukur proporsi instance yang benar diklasifikasikan sebagai positif dari semua instance yang diprediksi positif oleh model. Recall mengukur proporsi instance positif yang berhasil diidentifikasi oleh model dari keseluruhan instance positif yang sebenarnya. F1-score merupakan harmonik rata-rata dari precision dan recall, memberikan gambaran holistik tentang kinerja klasifikasi model. Selain itu, overall accuracy menunjukkan tingkat kesesuaian antara hasil prediksi dengan label sebenarnya dalam dataset secara keseluruhan. Analisis ini memberikan pemahaman yang mendalam tentang kemampuan model XGBoost dalam mengklasifikasikan setiap kelas serangan dapat dilihat pada gambar 2, serta keseluruhan kinerja model dalam mengatasi data yang kompleks dan tidak seimbang dalam dataset IoTID20.

Table 3. Analisis Performa XGBoost untuk Kelas Serangan pada Dataset IoTID20 Menggunakan Metrik Evaluasi

	DoS	MITM ARP Spoofing	Mirai	Normal	Scan
Precision	0.999915	0.959821	0.994300	0.996112	0.995329
Recall	0.998900	0.956365	0.997375	0.991017	0.983346
F1-Score	0.999407	0.958090	0.995835	0.993558	0.989301
Overall Accuracy	0.993161				



Gambar 2. Confusion Matrix

3.3. Pembahasan

Tabel 4 menyajikan perbandingan akurasi model klasifikasi multiclass yang diujicobakan pada dataset IoTID20. Model-model yang dievaluasi meliputi Deep Convolutional Neural Network (DCNN), Deep Learning dengan metode Margin-based Center Loss (Deep Learning-MCC), Single Hidden Layer Feed-Forward Neural Network (SLFN), Shallow Neural Networks (SNNs), Random Forest, serta Intersection Mathematical (IMF) dan Union Mathematical (UMF). Hasil akurasi model-model tersebut berkisar antara 88% hingga 100%. Model Shallow Neural Networks (SNNs) mencapai akurasi tertinggi, yaitu 100%, sementara model Decision Tree (Sub-Category) memiliki akurasi terendah sebesar 88%. Model yang diusulkan, yaitu FIR-XGBoost, menunjukkan akurasi sebesar 99.32%. Analisis ini memberikan pemahaman yang mendalam tentang performa relatif dari berbagai model klasifikasi multiclass dalam mengatasi dataset IoTID20, yang dapat menjadi dasar untuk pemilihan model yang optimal dalam aplikasi deteksi serangan pada jaringan IoT.

Table 4. Perbandingan Akurasi Model Klasifikasi Multiclass pada Dataset IoTID20

Author	Model	Akurasi
Ullah, Safi et al [25]	DCNN	98%
Y. Song [26]	Deep Learning-MCC	94%

R. Qaddoura [27]	Single Hidden Layer Feed-Forward Neural Network (SLFN)	98%
A. A. Alsulami [28]	Shallow Neural Networks (SNNs)	100%
P. Maniriho [29]	Random Forest (DoS, MITM, Scan)	99,96%
K. Albulayhi [30]	Intersection Mathematical (IMF) and Union Mathematical (UMF)	99.7% and 99,7%
I. Ullah [13]	Decision Tree (Sub-Category)	88%
Proposed Method	FIR-XGBoost	99.32%

4. KESIMPULAN

Dari hasil evaluasi, penggunaan algoritma FIR (Imbalance Ratio) dengan model XGBoost dalam klasifikasi dataset IoTID20 dapat disimpulkan bahwa Penggunaan algoritma FIR (Imbalance Ratio) dengan model XGBoost dalam klasifikasi dataset memberikan hasil yang sangat baik, dengan tingkat Accuracy 99.32%, Precision 99.32%, Recall 99.32%, dan F1 -Score 99.32%. Hal ini menunjukkan bahwa model XGBoost mampu mengatasi masalah ketidakseimbangan kelas dengan efektif, dan meningkatkan performa klasifikasi yang tinggi. Keberhasilan ini memberikan keyakinan bahwa kombinasi antara algoritma FIR dan XGBoost dapat menjadi solusi yang kuat dalam penanganan dataset dengan ketidakseimbangan kelas.

DAFTAR PUSTAKA

[1] T.-T.-H. Le, H. Kim, H. Kang, and H. Kim, "Classification and Explanation for Intrusion Detection System Based on Ensemble Trees and SHAP Method," *Sensors*, vol. 22, no. 3, 2022, doi: 10.3390/s22031154.

[2] U. Islam *et al.*, "Detection of Distributed Denial of Service (DDoS) Attacks in IOT Based Monitoring System of Banking Sector Using Machine Learning Models," *Sustain.*, vol. 14, no. 14, 2022, doi: 10.3390/su14148374.

[3] N. Dat-Thinh, H. Xuan-Ninh, and L. Kim-Hung, "MidSiot: A Multistage Intrusion Detection System for Internet of Things," *Wirel. Commun. Mob. Comput.*, vol. 2022, no. December 2017, 2022, doi: 10.1155/2022/9173291.

[4] N. Abughazaleh, R. Bin, M. Btish, and H. M., "DoS Attacks in IoT Systems and Proposed Solutions," *Int. J. Comput. Appl.*, vol. 176, no. 33, pp. 16–19, 2020, doi: 10.5120/ijca2020920397.

[5] M. Aljanabi, M. A. Ismail, and A. H. Ali, "Intrusion Detection Systems, Issues, Challenges, and Needs," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, pp. 560–571, Jan. 2021, doi: 10.2991/IJCIS.D.210105.001.

[6] P. Rieger, T. D. Nguyen, M. Miettinen, and A.-R. Sadeghi, "DeepSight: Mitigating Backdoor Attacks in Federated Learning Through Deep Model Inspection," Jan. 2022, doi: 10.14722/ndss.2022.23156.

[7] R. Yao, N. Wang, Z. Liu, P. Chen, and X. Sheng, "Intrusion detection system in the advanced metering infrastructure: A cross-layer feature-fusion CNN-LSTM-based approach," *Sensors (Switzerland)*, vol. 21, no. 2, pp. 1–17, 2021, doi: 10.3390/s21020626.

[8] A. Imran, "Performance Evaluation of Classification Algorithms for Intrusion Detection on NSL-KDD Using Rapid Miner," no. February, 2022.

[9] A. Alsaleh and W. Binsaeedan, "The influence of salp swarm algorithm-based feature selection on network anomaly intrusion detection," *IEEE Access*, vol. 9, pp. 112466–112477, 2021, doi: 10.1109/ACCESS.2021.3102095.

[10] S. Pokhrel, R. Abbas, and B. Aryal, "IoT Security: Botnet detection in IoT using Machine learning," Apr. 2021, [Online]. Available: http://arxiv.org/abs/2104.02231

[11] T. T. H. Le, Y. E. Oktian, and H. Kim, "XGBoost for Imbalanced Multiclass Classification-Based Industrial Internet of Things Intrusion Detection Systems," *Sustain. 2022, Vol. 14, Page 8707*, vol. 14, no. 14, p. 8707, Jul. 2022, doi: 10.3390/SU14148707.

[12] J. Al Amien, H. A. Ghani, N. I. M. Saleh, E. Ismanto, and R. Gunawan, "Intrusion detection system for imbalance ratio class using weighted XGBoost classifier," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 21, no. 5, pp. 1102–1112, 2023, doi: 10.12928/TELKOMNIKA.v21i5.24735.

[13] I. Ullah and Q. H. Mahmoud, "A Scheme for Generating a Dataset for Anomalous Activity Detection in IoT Networks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12109 LNAI, pp. 508–520, 2020, doi: 10.1007/978-3-030-47358-7_52.

[14] B. Li *et al.*, "A Systematic Review of Data-Driven Attack Detection Trends in IoT," *Sensors 2023, Vol. 23, Page 7191*, vol. 23, no. 16, p. 7191, Aug. 2023, doi: 10.3390/S23167191.

[15] A. A. Alfrhan, "SMOTE : Class Imbalance Problem In Intrusion Detection System," pp. 111–115, 2020.

[16] M. Son, S. Jung, J. Moon, and E. Hwang, "BCGAN-based over-sampling scheme for imbalanced data," *Proc. - 2020 IEEE Int. Conf. Big Data Smart Comput. BigComp 2020*, pp. 155–160, 2020, doi: 10.1109/BigComp48618.2020.00-83.

[17] J. Li *et al.*, "SMOTE-NaN-DE: Addressing the noisy and borderline examples problem in imbalanced classification by natural neighbors and differential evolution," *Knowledge-Based Syst.*, vol. 223, Jul. 2021, doi: 10.1016/j.knosys.2021.107056.

[18] Y. Fu, Y. Du, Z. Cao, Q. Li, and W. Xiang, "binary A Deep Learning Model for Network Intrusion Detection with Imbalanced Data," pp. 1–13, 2022.

[19] H. A. Ahmed, A. Hameed, and N. Z. Bawany, "Network intrusion detection using oversampling technique and machine learning algorithms," *PeerJ Comput. Sci.*, vol. 8, 2022, doi: 10.7717/PEERJ-CS.820.

[20] M. Koziarski, "Radial-Based Undersampling for imbalanced data classification," *Pattern Recognit.*, vol. 102, p. 107262, Jun. 2020, doi: 10.1016/J.PATCOG.2020.107262.

[21] N. Abdalgawad, A. Sajun, Y. Kaddoura, I. A. Zualkernan, and F. Aloul, "Generative Deep Learning to Detect Cyberattacks for the IoT-23 Dataset," *IEEE Access*, vol. 10, pp. 6430–6441, 2022, doi: 10.1109/ACCESS.2021.3140015.

[22] R. Zhu, Y. Guo, and J. H. Xue, "Adjusting the imbalance ratio by the dimensionality of imbalanced data," *Pattern Recognit. Lett.*, vol. 133, pp. 217–223, May 2020, doi: 10.1016/j.patrec.2020.03.004.

[23] K. Budholiya, S. K. Shrivastava, and V. Sharma, "An optimized XGBoost based diagnostic system for effective prediction of heart disease," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 7, pp. 4514–4523, Jul. 2022, doi: 10.1016/J.JKSUCI.2020.10.013.

[24] A. O. Alzahrani and M. J. F. Alenazi, "Designing a Network Intrusion Detection System Based on Machine Learning for Software Defined Networks," *Futur. Internet 2021, Vol. 13, Page 111*, vol. 13, no. 5, p. 111, Apr. 2021, doi: 10.3390/FI13050111.

[25] S. Ullah *et al.*, "A New Intrusion Detection System for the Internet of Things via Deep Convolutional Neural Network and Feature Engineering," *Sensors 2022, Vol. 22, Page 3607*, vol. 22, no. 10, p. 3607, May 2022, doi: 10.3390/S22103607.

[26] Y. Song, S. Hyun, and Y.-G. Cheong, "Analysis of autoencoders for network intrusion detection†," *Sensors*, vol. 21, no. 13, 2021, doi: 10.3390/s21134294.

[27] R. Qaddoura, A. M. Al-Zoubi, I. Almomani, and H. Faris, "A multi-stage classification approach for iot intrusion detection based on clustering with oversampling," *Appl. Sci.*, vol. 11, no. 7, 2021, doi: 10.3390/app11073022.

[28] A. A. Alsulami, Q. Abu Al-Haija, A. Tayeb, and A. Alqahtani, "An Intrusion Detection and Classification System for IoT Traffic with Improved Data Engineering," *Appl. Sci. 2022, Vol. 12, Page 12336*, vol. 12, no. 23, p. 12336, Dec. 2022, doi: 10.3390/APP122312336.

[29] P. Maniriho, E. Niyigaba, Z. Bizimana, V. Twiringiyimana, L. J. Mahoro, and T. Ahmad, "Anomaly-based Intrusion Detection Approach for IoT Networks Using Machine Learning," *CENIM 2020 - Proceeding Int. Conf. Comput. Eng. Network, Intell. Multimed. 2020*, no. Cenim, pp. 303–308, 2020, doi: 10.1109/CENIM51130.2020.9297958.

[30] K. Albulayhi, Q. A. Al-Haija, S. A. Alsuhibany, A. A. Jillepalli, M. Ashrafuzzaman, and F. T. Sheldon, "IoT Intrusion Detection Using Machine Learning with a Novel High Performing Feature Selection Method," *Appl. Sci.*, vol. 12, no. 10, 2022, doi: 10.3390/app12105015.