Infoman's : Jurnal Ilmu-ilmu Informatika dan Manajemen Vol.20 No.1 May (2026)
DOI: 10.33481/infomans.vxxixx

Benchmarking Machine Learning Models for Intrusion Detection in Higher Education Networks Using CIC-IDS2017

[1]Moch. Irwan Hermanto, [2]Galih, [3]Rudhi Wahyudi Febrianto, [4]Siti Nur
[1,2,4] Universitas Islam Nusantara
[3]Universitas Teknologi Bandung
email : [1]mochirwanh@uninus.ac.id, [2]galih@uninus.ac.id, [3]rudhiwahyudifebrianto@utb-univ.ac.id, [4]sitinuryendi@gmail.com

ABSTRACT
The security of university networks faces an increase in attack surfaces due to the high dependence on digital services. Machine learning-based Intrusion Detection System (IDS) is one of the approaches to recognize benign and malicious traffic patterns. This study analyzes the benchmark machine learning model in the CIC-IDS2017 dataset through a literature study approach and comparative analysis. The models compared include Decision Tree, Random Forest, Support Vector Machine, Naive Bayes, Artificial Neural Network, as well as additional benchmarks XGBoost, KNN, and Logistic Regression. Evaluation parameters include accuracy, precision, recall, F1-score, false positive, false negative, and class distribution. The benchmark results showed that Random Forest achieved an accuracy of 99.67% with precision, recall, and malicious class F1-scores of 99.90%, 99.67%, and 99.67%, respectively in one of the published configurations. Another comparative study reported that XGBoost and Random Forest achieved 99.96% accuracy in their experimental configurations. Analysis of the class distribution showed a strong imbalance so that accuracy was not sufficient to assess the effectiveness of the IDS. The study recommends macro-F1 evaluation, per-class recall, false negative, and testing on more representative data before the model is used in a college production environment.

Keywords *- CIC-IDS2017, cybersecurity, intrusion detection, machine learning*

1. Introduction

Transformasi digital pada perguruan tinggi meningkatkan ketergantungan terhadap jaringan komputer untuk layanan akademik, administrasi, pembelajaran, penelitian, dan komunikasi. Infrastruktur tersebut menghubungkan server, workstation, perangkat bergerak, access point, router, switch, firewall, aplikasi, dan layanan berbasis cloud. Peningkatan konektivitas secara bersamaan memperluas attack surface dan meningkatkan kebutuhan terhadap mekanisme deteksi aktivitas jaringan yang efektif.

Intrusion Detection System (IDS) merupakan salah satu komponen keamanan yang digunakan untuk mendeteksi aktivitas yang mengindikasikan intrusi. Pendekatan berbasis signature efektif terhadap pola serangan yang telah dikenal, tetapi memiliki keterbatasan terhadap variasi serangan baru. Machine learning menawarkan pendekatan berbasis pola dengan memanfaatkan karakteristik network flow sebagai fitur klasifikasi. Dataset CIC-IDS2017 dikembangkan untuk evaluasi IDS dan menyediakan traffic benign serta berbagai serangan dalam bentuk labeled network flows dan CSV untuk kebutuhan machine learning [1, 2].

Permasalahan utama dalam penerapan machine learning untuk IDS bukan hanya memperoleh accuracy tinggi, tetapi memastikan model mampu mendeteksi kelas serangan secara konsisten. Ketidakseimbangan kelas dapat menyebabkan accuracy dan weighted-average tinggi walaupun recall kelas minoritas rendah. False negative juga memiliki konsekuensi keamanan yang lebih serius karena

traffic serangan dapat dianggap normal. Oleh karena itu, evaluasi perlu mencakup precision, recall, F1-score, confusion matrix, dan analisis kesalahan [3]. Penelitian ini bertujuan menganalisis benchmark performa beberapa model machine learning pada CIC-IDS2017, mengidentifikasi implikasi class imbalance terhadap interpretasi hasil, dan merumuskan rekomendasi pemilihan model untuk konteks keamanan jaringan perguruan tinggi. Posisi penelitian adalah analisis benchmark berbasis literatur; angka numerik yang disajikan sebagai benchmark bukan hasil eksperimen mandiri yang belum dijalankan.

2. Research Method

Penelitian ini menggunakan Systematic Literature Review (SLR) yang dikombinasikan dengan sintesis komparatif kuantitatif. Metode dirancang untuk menjawab pertanyaan penelitian tanpa memperlakukan nilai benchmark dari konfigurasi eksperimen yang berbeda seolah-olah merupakan observasi dari satu eksperimen baru. Pembedaan ini penting karena perbedaan preprocessing, definisi kelas, pemilihan fitur, dan desain train-test dapat memengaruhi performa yang dilaporkan secara signifikan [3, 8, 10].

2.1 Pertanyaan Penelitian dan Ruang Lingkup

Ruang lingkup tinjauan dibatasi pada network-based intrusion detection menggunakan CIC-IDS2017/CICIDS2017 serta pendekatan ML atau deep learning untuk klasifikasi. Unit analisis utama adalah kombinasi model-metrik yang dilaporkan dalam penelitian terdahulu. Hasil binary dan multiclass dipertahankan sebagai konteks yang terpisah dan tidak digabungkan apabila pengaturan evaluasinya tidak kompatibel.

2.2 Strategi Pencarian

Strategi pencarian menggunakan kombinasi istilah "CIC-IDS2017", "CICIDS2017", "network intrusion detection", "machine learning", "deep learning", "Random Forest", "SVM", dan "XGBoost". Prioritas diberikan kepada artikel jurnal, prosiding konferensi, dokumentasi dataset resmi, dan publikasi NIST. Dokumentasi dataset menetapkan konteks CIC-IDS2017, sedangkan penelitian ilmiah menyediakan bukti perbandingan model [6, 12].

2.3 Kriteria Inklusi dan Eksklusi

Sebuah penelitian dimasukkan apabila (1) menggunakan CIC-IDS2017/CICIDS2017 sebagai dataset evaluasi atau benchmark utama, (2) membahas network intrusion detection, (3) menerapkan metode ML/deep learning, dan (4) melaporkan setidaknya satu metrik evaluasi kuantitatif. Penelitian dikeluarkan apabila bersifat konseptual, tidak menyediakan hasil terukur, tidak menjelaskan konteks dataset secara memadai, atau memberikan metrik yang tidak dapat diinterpretasikan secara konsisten.

2.4 Ekstraksi Data

Untuk setiap sumber yang memenuhi kriteria, skema ekstraksi mencatat model, jenis klasifikasi, accuracy, precision, recall, F1-score, support, dan catatan konfigurasi yang relevan. Support dipertahankan karena jumlah observasi setiap kelas diperlukan untuk menginterpretasikan metrik kelas minoritas. Nilai yang diekstraksi digunakan sebagai benchmark eksternal; nilai tersebut tidak dihitung ulang atau dipool secara statistik dari eksperimen yang tidak kompatibel.

DOI: 10.33481/infomans.vxxixx

2.5 Penilaian Kualitas dan Konsistensi

Kualitas bukti diperiksa melalui keterlacakan sumber, kejelasan dataset, kelengkapan metrik, dan konsistensi interpretasi. Perbedaan preprocessing, train-test split, jumlah kelas, serta klasifikasi binary dan multiclass dicatat sebagai bentuk heterogenitas. Oleh karena itu, analisis menekankan perbandingan deskriptif dan bukan meta-analysis formal terhadap effect size.

2.6 Sintesis Komparatif

Accuracy mengukur ketepatan keseluruhan, precision mengukur keandalan alert positif, recall mengukur proporsi instance malicious yang berhasil dideteksi, dan F1-score menyeimbangkan precision dan recall. Dalam cybersecurity, recall dan false negative mendapat perhatian khusus karena serangan yang tidak terdeteksi dapat menimbulkan kegagalan keamanan yang lebih serius dibandingkan tambahan alert yang tidak relevan [8, 10, 12, 16].

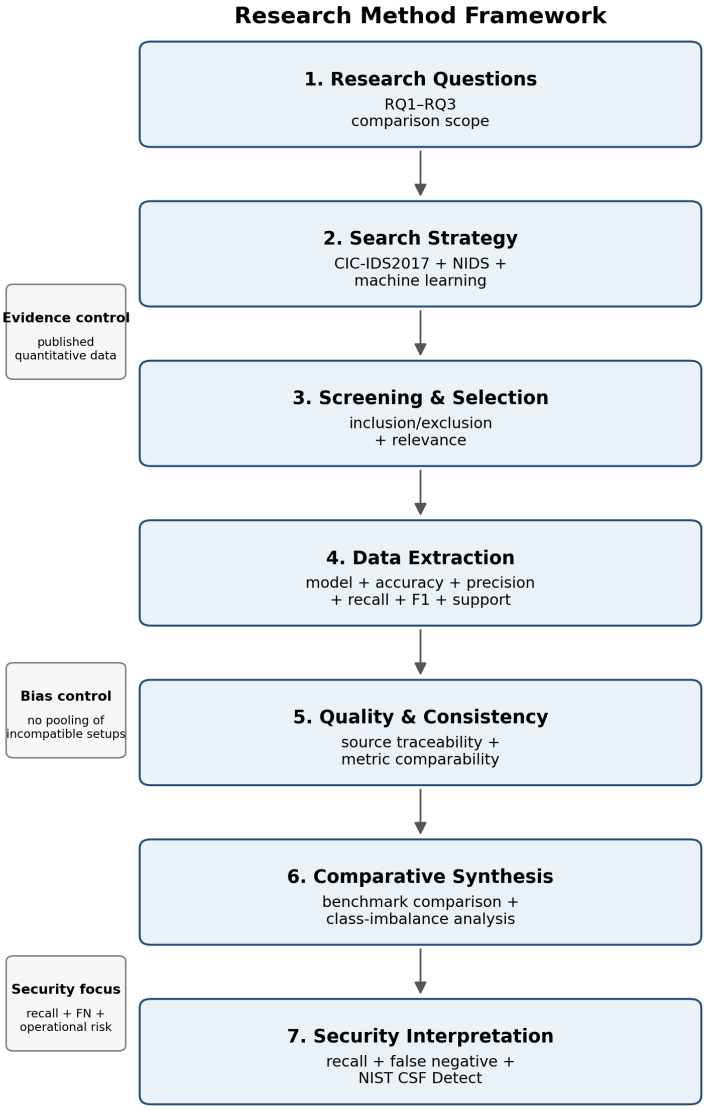

Figure 1. Research Method Framework

3. Result and Analysis

3.1 Benchmark Performance

Table 2. Benchmark performance on CIC-IDS2017

Model	Accuracy (%)	Precision malicious (%)	Recall malicious (%)	F1 malicious (%)
Decision Tree	99.59	99.87	99.59	99.31
Random Forest	99.67	99.90	99.67	99.67
SVM	96.00	99.27	96.10	96.10
Naive Bayes	72.96	93.60	64.18	78.29
ANN	95.49	98.31	92.64	95.39

Tabel 2 menunjukkan bahwa Random Forest memberikan accuracy tertinggi di antara model pada benchmark utama, yaitu 99,67%, dengan recall malicious 99,67% dan F1-score 99,67%. Decision Tree juga memberikan performa tinggi, sedangkan Naive Bayes menunjukkan penurunan yang lebih besar terutama pada recall malicious. Nilai tersebut merupakan benchmark dari konfigurasi penelitian terdahulu sehingga tidak dapat digeneralisasi sebagai hasil universal [11, 12].

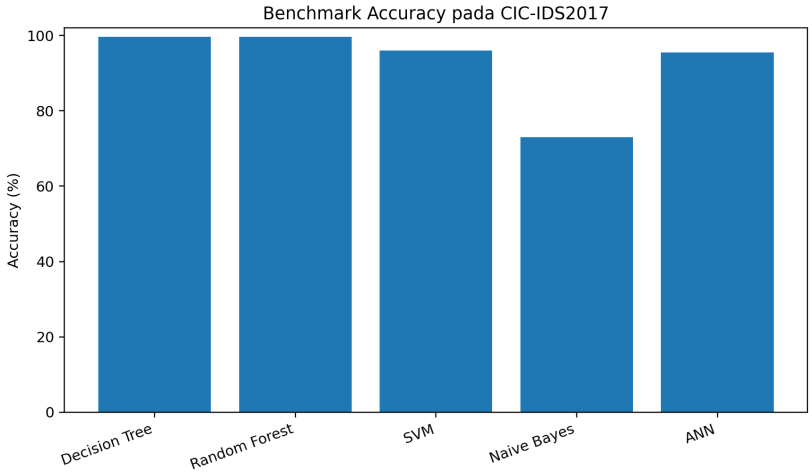

Figure 2. Benchmark accuracy model pada CIC-IDS2017

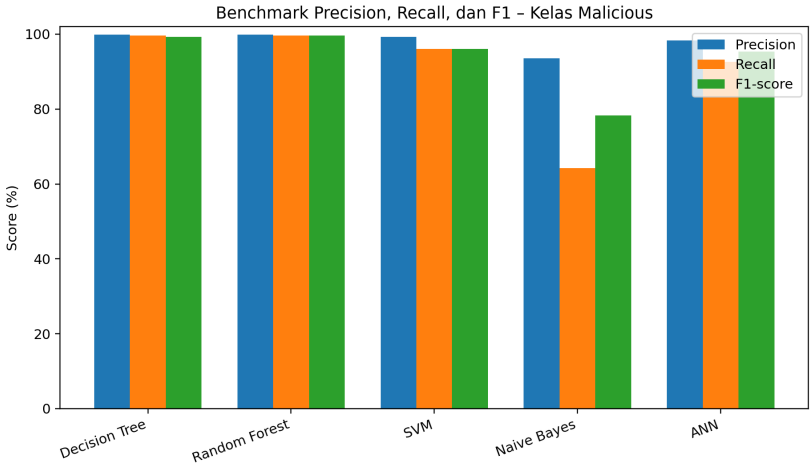

Figure 3. Benchmark precision, recall, dan F1-score kelas malicious

DOI: 10.33481/infomans.vxxixx

3.2 Comparative Benchmark

Table 3. Benchmark comparison from another CICIDS2017 study

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
XGBoost	99.96	99.99	100.00	100.00
KNN	98.58	98.43	98.59	98.43
Random Forest	99.96	99.98	99.99	100.00
SVM	75.36	72.23	99.95	82.34
Logistic Regression	94.41	89.44	98.69	94.85

Benchmark tambahan memperlihatkan XGBoost dan Random Forest memiliki accuracy 99,96% pada konfigurasi eksperimen yang dilaporkan. Perbedaan yang cukup besar pada SVM dibandingkan benchmark utama memperlihatkan pentingnya memperhatikan dataset subset, preprocessing, pembagian data, dan definisi kelas sebelum membandingkan angka antarpublikasi. [17]

3.3 Class Imbalance Analysis

Salah satu classification report Random Forest pada CICIDS2017 menunjukkan support yang sangat tidak seimbang. Kelas dengan support terbesar mencapai 9.066 sampel, sedangkan beberapa kelas hanya memiliki 4–18 sampel. Kondisi tersebut membuat nilai weighted-average dapat menutupi kelemahan pada kelas minoritas. [3]

Table 4. Class distribution in a published Random Forest benchmark

Class	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	9,066
1	0.50	0.25	0.33	8
2	1.00	1.00	1.00	467
3	1.00	0.95	0.98	42
4	0.99	1.00	1.00	939
5	1.00	0.93	0.97	30
6	1.00	1.00	1.00	23
7	1.00	1.00	1.00	28
8	1.00	1.00	1.00	672
9	1.00	0.94	0.97	18
10	0.75	0.60	0.67	10
11	0.25	0.25	0.25	4

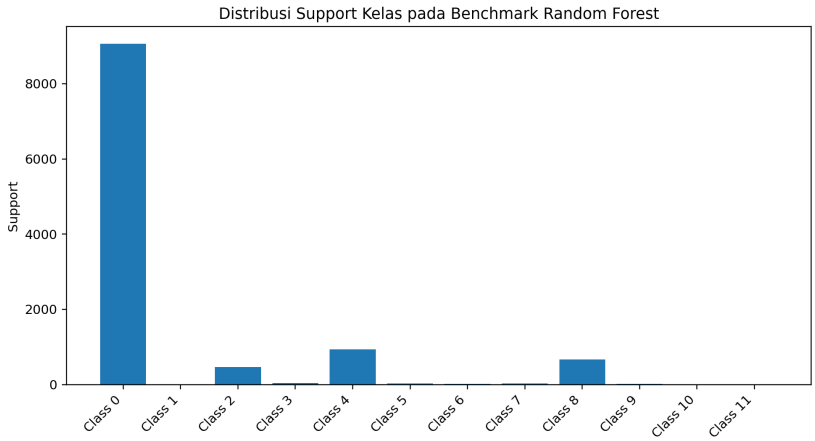

Figure 4. Support distribution pada benchmark Random Forest

Gambar 4 memperlihatkan bahwa kelas minoritas memiliki kontribusi sangat kecil terhadap total sampel. Misalnya, kelas 11 hanya memiliki empat observasi dan menghasilkan F1-score 0,25

pada benchmark tersebut. Temuan ini menjadi dasar bahwa evaluasi IDS tidak cukup menggunakan accuracy saja. Macro-F1 dan recall per kelas harus dipertimbangkan agar kemampuan deteksi serangan langka tidak tertutup oleh kelas mayoritas.

3.4 False Positive and False Negative

False positive terjadi ketika traffic benign diklasifikasikan sebagai serangan dan dapat meningkatkan jumlah alert. Jika terlalu banyak, operator keamanan berisiko mengalami alert fatigue. False negative terjadi ketika traffic serangan diklasifikasikan sebagai benign dan dapat menyebabkan serangan lolos dari deteksi. Dalam penelitian final, confusion matrix per model perlu digunakan untuk menghitung FP dan FN pada test set yang sama.

Table 5. Operational interpretation of detection errors

Error	Impact	Recommended analysis
False Positive	Alert overload	Threshold, precision, feature audit
False Negative	Attack evasion	Recall, class weighting, resampling
Minority-class error	Rare attack missed	Macro-F1 and per-class recall
Overfitting	Poor generalization	Cross-validation and temporal split
Data leakage	Optimistic performance	Pipeline and feature audit

4 Discussion

Hasil benchmark menunjukkan bahwa Random Forest dan XGBoost merupakan kandidat kuat untuk detection engine berbasis network flow. Namun, hasil antarpublikasi tidak identik. Perbedaan tersebut mengindikasikan bahwa model terbaik harus ditentukan pada konfigurasi data dan preprocessing yang sama. Penelitian lanjutan perlu menjalankan eksperimen mandiri dengan train-test split yang terdokumentasi, preprocessing tanpa data leakage, hyperparameter tuning pada training set, dan evaluasi macro-F1 [11-18].

Dalam konteks perguruan tinggi, model yang dipilih sebaiknya tidak hanya memiliki accuracy tinggi tetapi juga mampu menjaga recall serangan dan menghasilkan false positive yang dapat dikelola. Model yang memiliki sedikit false negative lebih sesuai untuk fungsi Detect, sedangkan precision diperlukan agar sistem tidak membanjiri administrator dengan alert.

Pemetaan terhadap NIST CSF 2.0 menempatkan prototype IDS pada fungsi Detect. Hasil deteksi dapat menjadi masukan bagi fungsi Respond melalui mekanisme alert dan investigasi. Integrasi lebih lanjut dengan SIEM, log management, atau SOC kampus dapat menjadi pengembangan berikutnya [4-10].

5 Conclusion

Analisis benchmark CIC-IDS2017 menunjukkan bahwa model ensemble berbasis tree, khususnya Random Forest dan XGBoost, memiliki performa kompetitif dan pada beberapa konfigurasi menghasilkan accuracy sekitar 99,67–99,96%. Namun, ketidakseimbangan kelas menunjukkan bahwa accuracy tinggi tidak selalu mencerminkan kemampuan deteksi seluruh jenis serangan. Beberapa kelas minoritas memiliki support sangat kecil dan F1-score rendah.

Dengan demikian, evaluasi IDS perlu menempatkan recall kelas serangan, macro-F1, false negative, dan confusion matrix sebagai indikator utama selain accuracy. Untuk penerapan pada jaringan perguruan tinggi, Random Forest atau XGBoost dapat dijadikan kandidat detection engine awal, tetapi pemilihan final harus divalidasi melalui eksperimen mandiri pada data yang representatif dan skenario pengujian yang dapat direproduksi. [8-16].

Penelitian selanjutnya disarankan menjalankan eksperimen langsung pada CIC-IDS2017 dan dataset kedua, menguji split berbasis waktu, melakukan feature selection dan hyperparameter tuning, serta mengembangkan prototype real-time yang terintegrasi dengan logging dan alert management.

DOI: 10.33481/infomans.vxxixx

References

[1] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," Proc. ICISSP, 2018.

[2] Canadian Institute for Cybersecurity, "Intrusion Detection Evaluation Dataset (CIC-IDS2017)," University of New Brunswick, dataset documentation.

[3] "Evaluation of Machine Learning Algorithms in Network-Based Intrusion Detection Using Progressive Dataset," Symmetry, 2023.

[4] National Institute of Standards and Technology, "The NIST Cybersecurity Framework (CSF) 2.0," NIST Cybersecurity White Paper 29, 2024.

[5] K. Rigopoulos, S. Quinn, C. Pascoe, A. Mahn, and D. Topper, "NIST Cybersecurity Framework 2.0: Resource & Overview Guide," NIST SP 1299, 2024.

[6] "Multi-Attack Intrusion Detection System for Software-Defined Internet of Things Network," comparative study using CICIDS2017.

[7] "Comparative Performance Evaluation of Machine Learning Algorithms for Cyber Intrusion Detection," Preprints, 2024.

[8] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in Proc. IEEE Symp. Security and Privacy, 2010, pp. 305–316, doi: 10.1109/SP.2010.25.

[9] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in Proc. IEEE Symp. Computational Intelligence for Security and Defense Applications (CISDA), 2009, pp. 1–6, doi: 10.1109/CISDA.2009.5356528.

[10] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, "A survey on machine learning techniques for cyber security in the last decade," IEEE Access, vol. 8, pp. 222310–222354, 2020, doi: 10.1109/ACCESS.2020.3041951.

[11] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," IEEE Access, vol. 7, pp. 41525–41550, 2019, doi: 10.1109/ACCESS.2019.2895334.

[12] Z. Ahmad, A. S. Khan, C. W. Shiang, J. Abdullah, and F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," Trans. Emerging Telecommun. Technol., vol. 32, no. 1, Art. no. e4150, 2021, doi: 10.1002/ett.4150.

[13] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study," J. Inf. Security Appl., vol. 50, Art. no. 102419, 2020, doi: 10.1016/j.jisa.2019.102419.

[14] S. Gamage and J. Samarabandu, "Deep learning methods in network intrusion detection: A survey and an objective comparison," J. Network and Computer Applications, vol. 169, Art. no. 102767, 2020, doi: 10.1016/j.jnca.2020.102767.

[15] Helmiawan, M. A., Herdiana, D., Firmansyah, E., Sholihah, A. I. N., Putri, S. W., & Septiana, S. (2025, September). Cybersecurity Awareness and Its Impact on the Association Between Technology Use and Adolescent Mental Health. In 2025 13th International Conference on Cyber and IT Service Management (CITSM) (pp. 1-6). IEEE.

[16] N. Shone, T. N. Nguyen, P. D. Vu, and Q. Shi, "A deep learning approach to network intrusion detection," IEEE Trans. Emerging Topics in Computational Intelligence, vol. 2, no. 1, pp. 41–50, 2018, doi: 10.1109/TETCI.2017.2772792.

[17] K. He, D. S. Kim, and M. R. Asghar, "Adversarial machine learning for network intrusion detection systems: A comprehensive survey," IEEE Commun. Surveys Tuts., vol. 25, no. 1, pp. 538–566, 2023, doi: 10.1109/COMST.2022.3233793.

[18] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.