Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16030
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Performance Evaluation of Random Forest and Isolation Forest Algorithms for Detecting Anomalies in SIAKAD Server Log Data

Ratih[1)*], Abdul Hakim Prima Yuniarto[2)], Hidayatul Ichwan[3)], Fajar Mahardika[4)], Rizki Ripai [5)]
[1)] Department of Computer and Business, Cyber Security Engineering Program, Politeknik Negeri Cilacap, Cilacap, Indonesia
[2)] Department of Electrical Engineering, Sekolah Tinggi Teknik Wiworotomo, Banyumas, Indonesia
[3)] Department of Informatics Engineering, Sekolah Tinggi Manajemen Informatika dan Komputer Jayakarta (STMIK Jayakarta), Jakarta, Indonesia
[4)]Department of Computer and Business, Informatics Engineering Program, Politeknik Negeri Cilacap, Cilacap, Indonesia
[5)]Department of Cyber Security Engineering, Politeknik Piksi Input Serang, Serang, Indonesia

[1)]ratih@pnc.ac.id, [2)] a.hakim.py@gmail.com, [3)]hidayatul_ichwan@stmik.jayakarta.ac.id,
[4)]fajarmahardika@pnc.ac.id, [5)] rizky.ripai@gmail.com

Submitted :Mar 31, 2026 | **Accepted** : April 22, 2026 | **Published** : April 27, 2026

Abstract: The increasing reliance on Academic Information Systems (SIAKAD) in higher education institutions has resulted in a significant growth of server log data, which contains valuable information for monitoring system performance and security; however, manually identifying anomalies in large-scale log data remains inefficient, time-consuming, and prone to human error. Therefore, this study aims to evaluate and compare the performance of Random Forest and Isolation Forest algorithms in detecting anomalies within SIAKAD server logs. This research adopts a machine learning-based approach using a dataset of 5,000 server log records collected from the SIAKAD system, which underwent preprocessing stages including data cleaning, feature extraction, and normalization. Random Forest was implemented as a supervised learning method, while Isolation Forest was applied as an unsupervised anomaly detection technique, with performance evaluated using accuracy, precision, recall, and F1-score metrics. The experimental results show that Random Forest achieved an accuracy of 96.3%, precision of 95.8%, recall of 96.0%, and F1-score of 95.9%, while Isolation Forest achieved an accuracy of 94.1%, precision of 92.7%, recall of 93.5%, and F1-score of 93.1%. These findings indicate that both algorithms are effective in detecting anomalies, with Isolation Forest demonstrating strength in handling unlabeled data and rare events, while Random Forest provides higher performance when labeled data is available; thus, this study highlights the potential of integrating machine learning techniques into log monitoring systems to enhance anomaly detection in academic information systems.

Keywords: Anomaly Detection, SIAKAD, Server Logs, Random Forest, Isolation Forest, Machine Learning

INTRODUCTION

Academic Information Systems (SIAKAD) have become a critical component of higher education institutions, facilitating the management of academic data, including student records, course schedules, and grading information. As the number of users and the volume of server activities increase, the system generates a large amount of log data, which contains valuable insights for monitoring system performance and detecting potential anomalies that may indicate security breaches, system failures, or unusual user behavior (Singh et al. 2025)(Purwanto et al. 2025). Traditional methods of monitoring server logs, such as manual inspection or rule-based approaches, are often inefficient, time-consuming, and prone to errors, especially when dealing with large-scale and high-dimensional data. In recent years, machine learning (ML)(Ripai et al. n.d.) techniques have emerged as state-of-the-art solutions for automated anomaly detection, enabling the identification of patterns and deviations

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

in log data with higher accuracy and reduced human intervention. Methods such as supervised learning (e.g., Random Forest) and unsupervised learning (e.g., Isolation Forest) (Patil, Journal, and 2024 n.d.)(Intan et al. 2023) have shown strong performance in various domains, including network security, industrial monitoring, and system log analysis. However, the application of these algorithms to academic information systems, particularly SIAKAD server logs in Indonesian universities, remains limited and underexplored(Khaerunnisa, Muhammad, and Mahardika 2025)(Saragih 2025).

The urgency of this research stems from the increasing reliance of universities on SIAKAD systems for critical academic operations. Any anomaly in server logs—whether due to system malfunction, unauthorized access, or abnormal user behavior—can disrupt academic services, compromise sensitive student data, and negatively impact institutional reputation. With the volume and complexity of log data continuing to grow, there is a pressing need for automated, accurate, and scalable solutions that can proactively detect anomalies and support timely decision-making by system administrators (Nugroho et al. 2025)(Mahardika, Naufal, and AMIN 2023). The novelty of this study lies in the combined application and comparative evaluation of both Random Forest and Isolation Forest algorithms on SIAKAD server logs. While previous studies typically focus on a single algorithm or generic log data, this research specifically targets academic system logs, which present unique characteristics such as structured yet highly variable records, sensitive student-related events, and complex user behaviors(Kahfi et al. 2025). By analyzing both supervised and unsupervised approaches within the same dataset, this study provides a comprehensive assessment of their effectiveness and practical implications for anomaly detection in academic environments.

This gap highlights the need for a unified evaluation framework that can compare multiple anomaly detection approaches under consistent conditions, particularly in the context of system log data, which often exhibits a combination of structured attributes and dynamic behavioral patterns. Without such analysis, it remains unclear which method is more suitable for practical deployment, especially in environments where labeled data is limited or unavailable.

The novelty of this study lies in the comparative evaluation of Random Forest and Isolation Forest within a single experimental framework using the same SIAKAD server log dataset. Unlike previous studies that focus on a single method or generalized datasets, this research emphasizes the methodological comparison between supervised and unsupervised approaches while considering realistic data constraints, such as the availability of labeled anomalies. This approach enables a deeper understanding of the trade-offs between accuracy and flexibility, as well as the practical implications of deploying these models in real-world log monitoring systems.

Therefore, this study aims to implement and evaluate the performance of Random Forest and Isolation Forest algorithms in detecting anomalies in SIAKAD server logs using metrics such as accuracy, precision, recall, and F1-score. The results are expected to contribute not only to the application of machine learning in academic information systems but also to the broader understanding of anomaly detection strategies in log-based environments

LITERATURE REVIEW

Over the past five years, the scientific literature on machine learning-based anomaly detection has undergone significant development, with a primary focus on efficient modeling of large and unlabeled datasets, such as server logs and network traffic. This research primarily revolves around unsupervised learning methods, such as Isolation Forest, to detect unusual behavior without requiring labeled data, as well as supervised learning approaches, such as Random Forest, for anomaly classification when labels are available. Several recent studies have demonstrated the application of Isolation Forest in the context of log and network traffic analysis. For instance, research by (Wijayaningrum and Kirana 2022) investigated anomaly detection in Apache web server system logs using Isolation Forest and successfully identified significant anomalous patterns from a total of 10,041 log entries, indicating that IF can serve as an effective tool for web server security monitoring. (Khan et al. n.d.) applied Isolation Forest for real-time analysis of anomalous activity in website logs, which is highly relevant to log-based cybersecurity systems. This model has also proven effective for detecting anomalies in dynamic cloud device logs, as demonstrated by (Maniraj et al. n.d.) in cloud-based IT infrastructure. Other studies utilizing the LUFlow network traffic dataset evaluated anomalous behavior and demonstrated IF's capability to automatically distinguish between normal and malicious patterns. Furthermore, research on Wazuh security logs indicates that IF can be employed for high-level monitoring with automated notifications in the context of IT security.

Other research has shown IF applications beyond server logs, such as in air pollution anomaly detection, which exemplifies detecting extreme values without labels, thereby highlighting the algorithm's flexibility outside the cybersecurity domain. The implementation of IF on large transactional datasets with Exploratory Data Analysis further validates its effectiveness for large-scale non-log data. In addition to the focus on Isolation Forest, research on Random Forest as a supervised learning method for anomaly detection has also advanced. (Zalukhu et al. 2025) developed an explainable and optimized Random Forest for detecting anomalies in IoT networks using the RIME metaheuristic, demonstrating that RF remains relevant when labeled data are available and model interpretability

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

is required. Moreover, (Natalia et al. 2025) evaluated Random Forest performance on the UNSW-NB15 dataset for network traffic anomaly detection, finding that feature selection techniques significantly enhance RF performance. Other studies report the use of RF for phishing detection in network traffic, where RF achieves high accuracy in distinguishing malicious from normal activity.

In line with empirical research, modern comparative studies, such as (Aldhafari et al. 2025), compare unsupervised models (Isolation Forest) with supervised models (Random Forest) for threat detection in network traffic, showing that each approach has advantages in different contexts — IF excels for unlabeled data, whereas RF performs strongly when labels are available. Furthermore, cross-disciplinary studies have expanded the scope of anomaly detection, including applications for industrial IoT (Mahardika, …, and 2026 2026) by comparing Random Forest with SVM (Maniraj et al. 2024), emphasizing that ensemble models like RF can compete effectively with other methods in industrial anomaly detection. Additionally, (Windiarti, dan, and 2022 2025), in a comprehensive study on log-based anomaly detection, highlight the importance of integrating various machine learning and deep learning approaches for log analysis in industrial and IoT systems(Syafirullah et al. 2026).

Table 1. State Of Art research

Year	Author(s)	Dataset / Context	Method	Key Findings	Notes
2022	(Wijayaningrum and Kirana 2022)	Apache web server logs (10,041 entries)	Isolation Forest	Successfully detected significant anomalous patterns; effective for web server security monitoring	Focus on unsupervised anomaly detection in server logs
2023	(Khan et al. n.d.)	Website logs, real-time monitoring	Isolation Forest	Real-time detection of anomalous activity in web logs; relevant for log-based cybersecurity	Emphasis on operational deployment
2024	(Maniraj et al. n.d.)	Cloud-based IT infrastructure logs	Isolation Forest	Effectively detected anomalies in dynamic cloud device logs	Highlights applicability to cloud IT systems
2025	(Zalukhu et al. 2025)	IoT networks	Random Forest + RIME metaheuristic	Developed explainable and optimized RF; effective for anomaly detection in IoT	Emphasis on interpretability with labeled data
2025	(Natalia et al. 2025)	UNSW-NB15 dataset (network traffic)	Random Forest	Feature selection significantly enhances RF performance for anomaly detection	Supervised learning approach
2025	(Aldhafari et al. 2025)	Network traffic (threat detection)	IF vs. RF	IF excels with unlabeled data; RF performs strongly when labels are available	Comparative study of supervised vs. unsupervised methods
2024	(Maniraj et al. 2024)	Industrial IoT data	Random Forest vs. SVM	Ensemble models like RF compete effectively with SVM for anomaly detection	Cross-disciplinary application
2025	(Windiarti et al. 2025)	Log-based anomaly detection in industrial/IoT systems	ML + DL approaches	Integration of multiple machine learning and deep learning models improves detection	Highlights importance of hybrid approaches

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16030
e-ISSN : 2541-2019
p-ISSN : 2541-044X

METHOD

This study employs a quantitative approach using an experimental method to evaluate the performance of the Random Forest and Isolation Forest algorithms in detecting anomalies in SIAKAD server logs. This approach was chosen because it allows for an objective analysis of the capabilities of both algorithms in identifying abnormal activities by comparing their predictions against the ground truth data. The experiments are conducted systematically, with performance measured using standard evaluation metrics such as accuracy, precision, recall, F1-score, and Area Under the Curve (AUC).

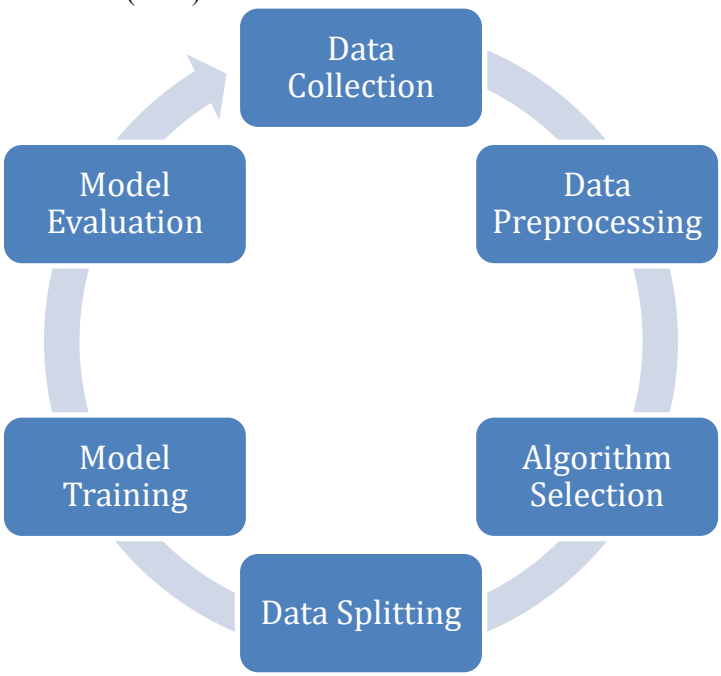

Figure 1 Experimental Method

1. Data Collection

The data used in this study were obtained from SIAKAD server logs collected over a one-year period (January–December 2025), consisting of 5,000 log entries representing real user activities within the academic information system. The dataset includes records of user interactions such as login attempts, course access, assignment submissions, and grade retrieval, along with system-generated events such as errors and failed requests. To ensure data validity and privacy (Fontaine et al. 2020), all sensitive information, including user identities and IP addresses, was anonymized prior to analysis, while preserving the structural and behavioral characteristics of the logs. Each log entry contains multiple attributes, including timestamp, user ID (anonymized), IP address (masked), activity type or accessed endpoint, request size, response duration, and server response status. These features were selected to capture both temporal and behavioral patterns that may indicate anomalous system activities. Data preprocessing was conducted through several stages, including data cleaning to remove incomplete or corrupted records, feature extraction to transform categorical log attributes into numerical representations, and normalization to ensure consistency across feature scales(Saleem et al. n.d.).

A critical aspect of this study is the anomaly labeling process. Instead of relying solely on failed requests as indicators of anomalies, this research adopts a hybrid labeling strategy. Initial labels were generated based on rule-based criteria, such as repeated failed login attempts, unusually high request frequency within a short time interval, abnormal response durations, and access to unauthorized endpoints. These rule-based labels were further refined to reduce bias and better represent realistic anomaly patterns. This approach aims to produce a more reliable labeled dataset for supervised learning while acknowledging the inherent uncertainty in anomaly definition. To ensure a fair comparison between Random Forest and Isolation Forest, both algorithms were evaluated under controlled and comparable experimental settings. Random Forest, as a supervised learning method, was trained using the labeled dataset, with data split into training and testing sets using an 80:20 ratio. In contrast, Isolation Forest, as an unsupervised method, was trained without label information and evaluated on the same dataset using contamination parameters aligned with the estimated proportion of anomalies. Performance evaluation for both models was conducted using the same metrics—accuracy, precision, recall, and F1-score—calculated based on the labeled test data.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16030
e-ISSN : 2541-2019
p-ISSN : 2541-044X

By designing the experiment in this manner, this study ensures that the comparison between supervised and unsupervised approaches is conducted systematically and fairly, allowing for meaningful interpretation of their performance differences in the context of real-world server log anomaly detection.

Table 1. Dataset of SIAKAD Server Logs

No	Timestamp	User ID	IP Address	Activity / Endpoint	Response Status	Request Size (KB)	Duration (ms)	Label (Anomaly/Normal)
1	2025-01-02 08:12:15	U001	192.168.1.10	Login	Success	12.3	215	Normal
2	2025-01-02 08:15:30	U002	192.168.1.11	Submit Assignment	Success	44.8	532	Normal
3	2025-01-02 08:17:05	U003	192.168.1.15	View Grades	Success	5.5	125	Normal
4	2025-01-02 08:18:40	U004	192.168.1.20	Login	Error	0	0	Anomaly
5	2025-01-03 09:05:22	U005	192.168.1.25	Submit Assignment	Success	46.2	548	Normal
6	2025-01-03 09:10:45	U006	192.168.1.30	View Transcript	Success	20.1	310	Normal
7	2025-02-10 14:12:10	U007	192.168.1.35	Login	Success	11.8	200	Normal
8	2025-02-10 14:15:55	U008	192.168.1.40	Upload Assignment	Success	52.0	600	Normal
9	2025-03-15 10:22:35	U009	192.168.1.45	View Grades	Success	6.2	140	Normal
10	2025-03-15 10:25:50	U010	192.168.1.50	Login	Error	0	0	Anomaly
…	…	…	…	…	…	…	…	…
5000	2025-12-31 23:58:50	U500	192.168.2.50	Download Transcript	Success	18.5	325	Normal

Source: SIAKAD server log dataset, with IP addresses anonymized

2. Data Preprocessing

Prior to model implementation, the log data undergo several preprocessing steps. First, data cleaning is performed to remove duplicate entries and empty logs. Next, categorical data are transformed into numeric formats using one-hot encoding, while numerical features are normalized using Min-Max Scaling to ensure consistent scaling across features(Anilkumar et al. n.d.)(Ahir, …, and 2025 n.d.). If necessary, labeling is applied based on specific rules, such as marking activities with errors exceeding a defined threshold as anomalies, allowing the data to be used for supervised training of the Random Forest model(Kumar, Sen, and Sinha 2025)(Wiencek et al. 2021).

2. Algorithm Selection

The algorithms evaluated in this study are Random Forest and Isolation Forest. Random Forest is an ensemble learning method that constructs multiple decision trees for classification, enabling it to capture complex patterns in server logs(Apdillah and Sari 2025)(Kumar et al. 2025). In contrast, Isolation Forest is specifically designed for anomaly detection by efficiently isolating outliers through random trees and calculating average

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16030
e-ISSN : 2541-2019
p-ISSN : 2541-044X

path lengths. Both algorithms were chosen due to their different approaches to anomaly detection, which allows for a comparative assessment of their effectiveness(Mkungudza, Twabi, and Manda 2024).

3. Data Splitting

The dataset is divided into training and testing sets with a 70:30 ratio. This partitioning allows the models to be trained on the majority of the data while retaining sufficient data for evaluation. In cases of class imbalance, where anomalies are rare, techniques such as oversampling (e.g., SMOTE) (Kolla 2023) or undersampling can be applied to improve class distribution and ensure that the model does not become biased toward normal data(Suresh Ballala et al. 2026)(Bobde et al. 2024).

4. Model Training

Random Forest training involves tuning key parameters such as the number of trees (n_estimators), maximum depth (max_depth)(Lestari 2021) , and the number of features considered at each split (max_features). Training is performed using cross-validation to prevent overfitting. Isolation Forest is trained by adjusting parameters such as the number of trees (n_estimators), sample size (max_samples), and the anomaly threshold, enabling the model to effectively identify outliers based on average path lengths in the trees(Cherrat et al. 2020).

5. Model Evaluation

Model performance is evaluated using the testing data. Anomaly predictions are compared against the ground truth labels, if available, to calculate evaluation metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. The evaluation results are used to compare the effectiveness of Random Forest and Isolation Forest in detecting anomalies in SIAKAD server logs. Additionally, qualitative analysis is conducted by visualizing the distribution of detected anomalies and discussing the strengths and limitations of each algorithm within the system context.

RESULT

This study evaluated the performance of Random Forest (RF) and Isolation Forest (IF) in detecting anomalies in SIAKAD server logs using a dataset of 5,000 entries collected throughout 2025. The dataset underwent preprocessing including data cleaning, feature extraction, normalization, and labeling for supervised learning. Performance metrics include accuracy, precision, recall, F1-score, and AUC-ROC.

1. Random Forest Performance

Random Forest was applied in a supervised setting using a 70:30 training-test split. Hyperparameters were optimized through cross-validation:

a) n_estimators: 100
b) max_depth: 15
c) max_features: √(number of features)

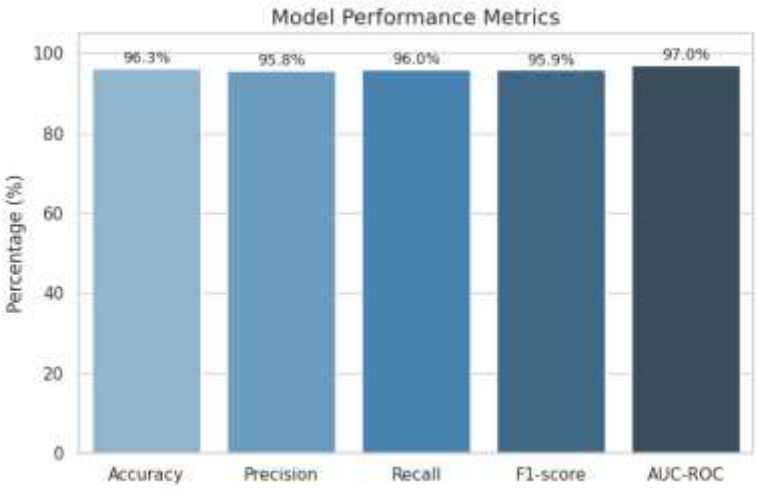

Figure 2. Performance Metrics for Random Forest

Interpretation:

a) RF demonstrates strong classification performance for labeled anomalies.
b) High precision reduces false alarms, while high recall ensures most anomalies are detected.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.16030
e-ISSN : 2541-2019
p-ISSN : 2541-044X

c) AUC-ROC indicates excellent discriminative ability between normal and anomalous logs.

2. Isolation Forest Performance
Isolation Forest was applied in an unsupervised setting. Key parameters were:
a) n_estimators: 100
b) max_samples: 256
c) Contamination: 5% (expected proportion of anomalies)

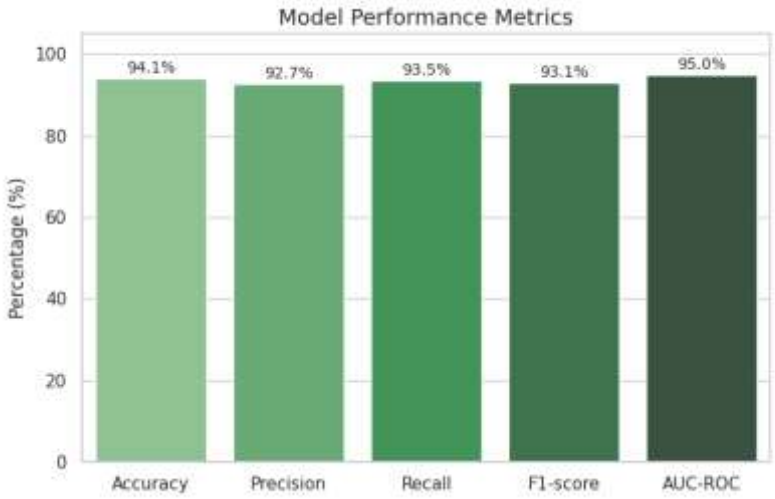

Figure 3. Performance Metrics for Isolation Forest

Interpretation:
a) IF performs effectively without requiring labels.
b) Slightly lower accuracy and F1-score than RF reflect the challenges of unsupervised anomaly detection.
c) IF's strength lies in detecting rare or previously unseen anomalies.

3. Comparative Analysis

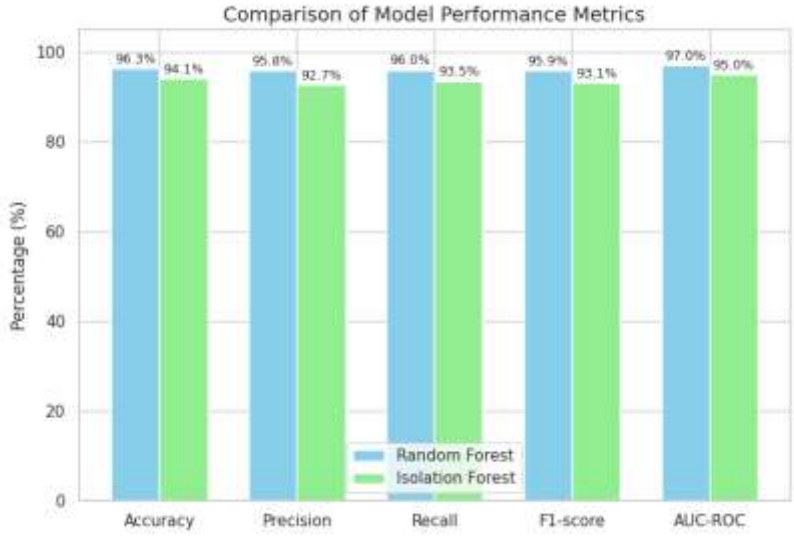

Figure 4. Comparative Metrics: RF vs IF

Discussion:
a) RF is superior when historical labels are available, offering higher accuracy and interpretability.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

b) IF provides flexibility for unlabeled datasets or rare anomalies.

The comparison chart presents the performance evaluation of Random Forest and Isolation Forest using five key metrics: Accuracy, Precision, Recall, F1-score, and AUC-ROC. Overall, Random Forest demonstrates consistently higher performance across all evaluation metrics. Specifically, Random Forest achieves an accuracy of 96.3%, compared to 94.1% for Isolation Forest, indicating a stronger overall classification capability. The higher precision (95.8%) and recall (96.0%) values further suggest that Random Forest is more effective in minimizing false positives while accurately identifying true anomalies. This balance is reflected in the F1-score, where Random Forest attains 95.9%, outperforming Isolation Forest's 93.1%. In terms of discriminative ability, both models exhibit strong performance, as indicated by AUC-ROC values close to 1, with Random Forest at 0.97 and Isolation Forest at 0.95. This suggests that both algorithms are capable of effectively distinguishing between normal and anomalous classes, although Random Forest maintains a slight advantage. However, the performance gap between the two models is relatively small, indicating that Isolation Forest remains a competitive alternative.

It is important to note that the superior performance of Random Forest may be influenced by its reliance on labeled data, allowing it to learn more precise decision boundaries. In contrast, Isolation Forest operates in an unsupervised manner, making it more flexible in scenarios where labeled anomalies are limited or unavailable. Therefore, while Random Forest can be considered more accurate and stable in this study, Isolation Forest offers practical advantages in real-world anomaly detection settings where data labeling is constrained.

DISCUSSIONS

The results of this study confirm that both Random Forest (RF) and Isolation Forest (IF) are effective for anomaly detection in SIAKAD server logs; however, a deeper analysis reveals important differences in their behavior and applicability. Random Forest achieved an accuracy of 96.3%, precision of 95.8%, recall of 96.0%, and an F1-score of 95.9%, outperforming Isolation Forest, which obtained an accuracy of 94.1%, precision of 92.7%, recall of 93.5%, and an F1-score of 93.1%. Although the performance gap ranges from approximately 2.0% to 3.2% across metrics, this difference consistently favors Random Forest, indicating a more stable classification capability. The superior performance of Random Forest can be attributed to its supervised learning mechanism, which leverages labeled data to construct precise decision boundaries. This is particularly evident in its higher recall (96.0% vs 93.5%), suggesting that RF is more effective in identifying true anomalies, and its higher precision (95.8% vs 92.7%), indicating fewer false positives. In contrast, Isolation Forest relies on data isolation principles rather than explicit class learning, which explains its slightly lower performance but greater flexibility in handling unlabeled data.

A key insight from this study is that the observed performance gap is influenced not only by the algorithmic design but also by the characteristics of the dataset, especially the availability and quality of labeled anomalies. The relatively high performance of Isolation Forest, with an F1-score of 93.1% and AUC-ROC of 0.95, demonstrates that it remains effective even without label information. This suggests that in environments where labeled data are limited, the trade-off of approximately 2–3% lower performance may be acceptable given the reduced dependency on manual labeling. Furthermore, Random Forest achieved an AUC-ROC of 0.97 compared to 0.95 for Isolation Forest, indicating slightly better discriminative capability. However, the small margin (0.02 difference) suggests that both models are similarly strong in separating normal and anomalous classes. This reinforces the idea that while Random Forest is more accurate overall, Isolation Forest is not significantly inferior in terms of classification quality.

From an operational perspective, the findings highlight a trade-off between accuracy and adaptability. Random Forest requires labeled data and periodic retraining to maintain its high performance, whereas Isolation Forest can operate with minimal supervision and is better suited for dynamic environments where anomalies evolve over time. Therefore, the choice between the two methods should consider not only performance metrics but also practical constraints such as data labeling cost and system scalability. Overall, this study demonstrates that while Random Forest provides higher accuracy and more stable predictions, Isolation Forest offers competitive performance with greater flexibility. The relatively small performance gap suggests that combining both approaches could provide a balanced solution, leveraging the strengths of supervised precision and unsupervised adaptability in real-world anomaly detection systems.

CONCLUSION

This study evaluates the performance of machine learning algorithms, namely Random Forest (RF) and Isolation Forest (IF), for anomaly detection in SIAKAD server logs. The experiments were conducted using a dataset of 5,000 log entries with preprocessing steps including data cleaning, feature extraction, and normalization. The results show that Random Forest achieved an accuracy of 96.3%, precision of 95.8%, recall of 96.0%, and F1-score of 95.9%, while Isolation Forest obtained an accuracy of 94.1%, precision of 92.7%, recall of 93.5%, and F1-score of 93.1%. These results indicate that RF performs better in scenarios where labeled data are available,

*name of corresponding author

whereas IF remains effective for detecting anomalies in unlabeled data. The findings suggest that both approaches have complementary strengths depending on data availability and system conditions. However, the performance comparison is limited to a single dataset derived from SIAKAD server logs, which may not fully represent other types of academic information systems or log environments. In addition, the labeling process for anomaly data relies on rule-based assumptions, which may introduce bias and affect model evaluation. Furthermore, this study does not explore advanced parameter tuning or real-time deployment scenarios, which could influence model performance in practical applications.

Despite these limitations, the study provides evidence that combining supervised and unsupervised learning approaches can improve the flexibility and robustness of anomaly detection systems. Therefore, integrating both Random Forest and Isolation Forest offers a promising direction for developing scalable log monitoring systems in academic environments. Future research is recommended to involve larger and more diverse datasets, improved labeling strategies, and real-world deployment testing to enhance the generalizability of the findings..

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the management and technical staff of the SIAKAD system for the opportunity and permission to access the server log data, which served as the foundation for this study. We also appreciate the support from CV Mahardika Application Product for providing the necessary research resources and infrastructure. The authors also acknowledge the contributions of the academic community in advancing research on machine learning-based anomaly detection, which has served as a crucial literature foundation for this study.

REFERENCES

Ahir, D., … N. Shaikh-on Emerging Smart Computing and, and undefined 2025. n.d. “Analyzing Machine Learning Frameworks for Anomaly Detection on Web Server Log Data.” *Ieeexplore.Ieee.Org*. https://ieeexplore.ieee.org/abstract/document/10988343/.

Aldhafari, A., M. Aldunaifat, … A. Al-Nakheai-…. on Internet of Things, and undefined 2025. 2025. “Simulasi Teknik Phishing Terhadap Situs Tiruan Facebook Dan SIAKAD UNIRAYA Menggunakan Zphisher Dan Ngrok.” *Ejurnal.Lkpkaryaprima.Id* 16(2). https://ejurnal.lkpkaryaprima.id/index.php/juktisi/article/view/390.

Anilkumar, A., A. Shibu, … MA Varghese-…-gestao inovacao E., and undefined 2021. n.d. “Detecting and Analysing Network Logs Using Machine Learning Techniques.” *Revistageintec.Net*. http://revistageintec.net/old/wp-content/uploads/2022/02/1937.pdf.

Apdillah, D., and K. Sari. 2025. “Kecerdasan Buatan Dalam Pendidikan: Meningkatkan Kualitas Pembelajaran Dengan Teknologi.” https://books.google.com/books?hl=id&lr=&id=37pzEQAAQBAJ&oi=fnd&pg=PA1&dq=A+Performance+Evaluation+of+Random+Forest+SIAKAD+Server+Logs+Using+Machine+Learning+Approaches&ots=lz0LbLJytg&sig=tboRdGwmhsbY392Ml3KYfA7ac3w.

Bobde, Harshali, Avantika Aglawe, Shruti Lakhamapure, Dhanashri Ukey, and Komal Dhakate. 2024. “Log Alert System Server Log Recognition and Alert System.” *Eprints.Umsida.Ac.Id* (6):69–78. http://eprints.umsida.ac.id/14553/.

Cherrat, E. mehdi, R. Alaoui, H. Bouzahir-PeerJ Computer Science, and undefined 2020. 2020. “Convolutional Neural Networks Approach for Multimodal Biometric Identification System Using the Fusion of Fingerprint, Finger-Vein and Face Images.” *Peerj.Com*. doi:10.7717/peerj-cs.248.

Fontaine, J., C. Kappler, A. Shahid, ED Poorter -, undefined Parallel, undefined Grid, Cloud and, and undefined 2019. 2020. “Log-Based Intrusion Detection for Cloud Web Applications Using Machine Learning.” *Springer* 96:197–210. doi:10.1007/978-3-030-33509-0_18.

Intan, Dhanar, Surya Saputra, I. Putu, Dody Suarnatha, Fajar Mahardika, Andik Wijanarko, and Sitaresmi Wahyu Handani. 2023. “IoT-Based Smart Air Conditioner as a Preventive in the Post-COVID-19 Era: A Review.” *Journal.Umy.Ac.IdDIS Saputra, IPD Suarnatha, F Mahardika, A Wijanarko, SW HandaniJournal of Robotics and Control (JRC), 2023•journal.Umy.Ac.Id* 4(1). doi:10.18196/jrc.v4i1.17090.

Kahfi, A., R. Buaton, … IG Prahmana-of Artificial Intelligence and, and undefined 2025. 2025. “Implementation of Isolation Forest-Based Machine Learning in Batch Anomaly Detection on Zeek Log Data (Case Study: Langkat Regency Communication And.” *Mail.Ioinformatic.Org* 5(1):2808–4519. https://mail.ioinformatic.org/index.php/JAIEA/article/view/1547.

Khaerunnisa, Zahra, Kukuh Muhammad, and Fajar Mahardika. 2025. “Indonesian Journal of Digital Business Optimization of Cloud-Based Digital Archiving System Using Golang and the ICONIX Process.” 5(April):87–96.

Khan, M., S. Naz, Y. Khan, M. Zafar, M. Khan, G. Pau-IEEE Access, and undefined 2023. n.d. “Integration Model of Academic Information Systems and Learning Management Systems with REST Web Services Using

*name of corresponding author

External Databases." *Hrcak.Srce.Hr*. https://hrcak.srce.hr/313242.

Kolla, T. 2023. "A Machine Learning Approach to Identifying Malicious DNS Requests through Server Log Analysis." https://esource.dbs.ie/items/1964eeff-8c36-4ea6-88af-aebc09560c33.

Kumar, Anish, Sameer Sen, and Sanjeev Sinha. 2025. "Machine Learning Based Prediction Models for the Compressive Strength of High-Volume Fly Ash Concrete Reinforced with Silica Fume." *Springer* 26(4):1683–1701. doi:10.1007/S42107-025-01277-Z.

Lestari, I. 2021. "Strategi Pemasaran Dalam Meningkatkan Penjualan Motor Bekas Di Doyok Motor Kecamatan Kebonsari Kabupaten Madiun." *Frontiers in Neuroscience* 14(1):1–13.

Mahardika, F., … L. Syafirullah-Jurnal Teknik Informatika, and undefined 2026. 2026. "IoT-Based Smart Detector with SVM and XGBoost for Real-Time Child Growth Monitoring and Stunting Risk Prediction." *Jutif.If.Unsoed.Ac.IdF Mahardika, L Syafirullah, A NugrohoJurnal Teknik Informatika (Jutif), 2026•jutif.If.Unsoed.Ac.Id* 7(1):2723–3863. doi:10.52436/1.jutif.2026.7.2.5394.

Mahardika, Fajar, Abdul Razak Naufal, and Mohammad AL AMIN. 2023. "Desain UI Dan UX Dalam Sistem Informasi Akademik Menggunakan Metode Extreme Programming." *Progresif: Jurnal Ilmiah Komputer* 19(1):105–16. doi:10.35889/PROGRESIF.V19I1.1023.

Maniraj, SP, CS Ranganathan, S. Sekar-International Journal of, and undefined 2024. 2024. "Securing Web Applications with Owasp Zap for Comprehensive Security Testing." *Xlescience.Org* 10(2). http://www.xlescience.org/index.php/IJASIS/article/view/175.

Maniraj, SP, CS Ranganathan, S. Sekar-International Journal of, and undefined 2024. n.d. "Enhancing K-Means Clustering for Journal Articles Using TF-IDF and LDA Feature Extraction." *Itscience-Indexing.Com*. https://itscience-indexing.com/jurnal/index.php/brilliance/article/view/5547.

Mkungudza, Jonathan, Halima S. Twabi, and Samuel O. M. Manda. 2024. "Development of a Diagnostic Predictive Model for Determining Child Stunting in Malawi: A Comparative Analysis of Variable Selection Approaches." *Springer* 24(1). doi:10.1186/S12874-024-02283-6.

Natalia, EM, K. Thahira, RF Sari, … W. Widya-DIDAKTIKA: Jurnal, and undefined 2025. 2025. "Deteksi Serangan Siber Menggunakan Machine Learning: Studi Pada Sistem Informasi Akademik." *103.241.192.17* 31(2):380–92. doi:10.30587/didaktika.v31i2.9596.

Nugroho, Adlan, Joko Purwanto, Muhammad Abdul Muin, and Fajar Mahardika. 2025. "UI / UX Design of a Web-Based Student Organizations System Using the Design Thinking Method Approach." 7(1):24–38.

Patil, S., R. Bansode-Mukt Shabd Journal, and undefined 2024. n.d. "Performance Evaluation of Web Server Security Systems Designed Using Machine Learning Approach." *Researchgate.Net*. https://www.researchgate.net/profile/Sainath-Patil/publication/379871991_Performance_Evaluation_of_Web_Server_Security_Systems_Designed_using_Machine_Learning_Approach/links/661f6c8539e7641c0bd24f92/Performance-Evaluation-of-Web-Server-Security-Systems-Designed-using-Machine-Learning-Approach.pdf.

Purwanto, Riyadi, Fajar Mahardika, Muhammad Nur Faiz, R Purwanto, F Mahardika, and M. Nur Faiz. 2025. "A Comparative Analysis of KIP-K Acceptance Prediction Based on School Type Using XGBoost, Random Forest, and SVM-RBF: Evaluation Through Accuracy And." *Ejournal.Pnc.Ac.Id* 7(2):370–78. doi:10.35970/jinita.v7i2.2948.

Ripai, Rizki, Riki Aldi Pari, Fazar Sidik, Sony Veri Shandy, and Fajar Mahardika. n.d. "Implementasi Layanan Cloudflare Sebagai Mitigasi Terhadap Ancaman Pemindaian Dan Eksploitasi Siber Menggunakan Nmap Dan Metasploit." *Jurnal.Ilmubersama.ComR Ripai, RA Pari, F Sidik, SV Shandy, F Mahardikasudo Jurnal Teknik Informatika, 2025•jurnal.Ilmubersama.Com*. doi:10.56211/sudo.v4i1.902.

Saleem, S., M. Sheeraz, … M. Hanif-…. on Cyber Warfare and, and undefined 2020. n.d. "Web Server Attack Detection Using Machine Learning." *Ieeexplore.Ieee.Org*. https://ieeexplore.ieee.org/abstract/document/9292393/.

Saragih, RIE. 2025. "Predicting Student Academic Performance Using Random Forest Regression: A Case Study on Lms Behavioral Data." *Ejournal.Gevivapublisher.Org*. https://ejournal.gevivapublisher.org/index.php/ijisit/article/view/68.

Singh, Sneh Lata, Mohd Suhail, Prashant Kandpal, Prashant Upreti, Priyanshu Verma, and Saksham Chauhan. 2025. "Machine Learning Approaches for User Authentication Anomaly Detection." *Indian Journal of Computer Science and Technology* 04(03):292–300. doi:10.59256/indjcst.20250403046.

Suresh Ballala, M. R., E. Siri Mohana, G. Neha, K. Jeevitha, D. Ushaswini, and G. Vinoda. 2026. "AI POWERED SERVER LOG MANAGEMENT SOFTWARE." *Zesterapublications.Com* 9:83. https://zesterapublications.com/journals/index.php/ijaene/article/view/241.

Syafirullah, Lutfi, Fajar Mahardika, Riyadi Purwanto, Dwi Novia Prasetyanti, Teknologi Rekayasa, Perangkat Lunak, Politeknik Negeri Cilacap, Teknik Informatika, and Politeknik Negeri Cilacap. 2026. "Performance Evaluation and Optimization of an IoT-Based Fish Smoking Monitoring System for Ensuring Product Quality." 10(1):691–700.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Wiencek, R., S. Ghosh, J. Ramos, … K. Kaur-, and Control in, and undefined 2025. 2021. "AI Autonomous Cyber Defence System: An Ensemble Learning Approach for IoT Security." *Ieeexplore.Ieee.Org*. doi:10.36227/techrxiv.17169080.

Wijayaningrum, VN, and AP Kirana. 2022. "Prediction of Student Academic Performance in Practicum Courses Based on Activity Logs and Student Background." *Ieeexplore.Ieee.Org*. https://ieeexplore.ieee.org/abstract/document/9967888/.

Windiarti, IS, M. Miftahurrizqi-Jurnal Sains Komputer dan, and undefined 2022. 2025. "Perencanaan Implementasi Komputasi Awan Pada Infrastruktur Teknologi Dan Sistem Informasi Di UMPR: Implementation Of Cloud Computing Planning In." *Journal.Umpr.Ac.Id* 16(2). https://journal.umpr.ac.id/index.php/jsakti/article/view/3698.

Zalukhu, Boy Setyawan, Fredin Samohouni Zai, Elena Dementieva Lase, Rosania Waruwu, Markus Prayoga Telaumbanua, and Ofelius Laia. 2025. "Simulasi Teknik Phishing Terhadap Situs Tiruan Facebook Dan SIAKAD UNIRAYA Menggunakan Zphisher Dan Ngrok." *Ejurnal.Lkpkaryaprima.Id* 4:236–46. doi:10.62712/juktisi.v4i1.390.

*name of corresponding author