Journal Article

International Journal for Science Review

Optimization of Machine Learning Algorithms for Enhancing Cybersecurity in Cloud Computing Environments

Badie Uddin

Esa Unggul University, Jalan Arjuna Utara No. 9, Kebon Jeruk, West Jakarta 11510, Indonesia.

*Corresponding author: badie.uddin@esaunggul.ac.id

ABSTRACT

Advances in cloud computing technology have driven digital transformation in various sectors, but on the other hand, it has also posed serious challenges in terms of cybersecurity. Threats such as Distributed Denial of Service (DDoS) attacks, data leaks, and unauthorized access are increasingly complex and difficult to detect with traditional approaches. This article aims to examine in depth how machine learning algorithm optimization can increase the effectiveness of cybersecurity systems in cloud computing environments. This research uses a qualitative approach with a literature study method or library research that focuses on the analysis of various scientific publications, technical reports, and current case studies. The results of the study show that algorithms such as Random Forest, Support Vector Machine (SVM), and Deep Learning have significant potential in detecting and responding to threats in real-time. Optimization is carried out through improving the quality of the dataset, tuning model parameters, and integration with cloud-based intrusion detection systems. In addition, the use of ensemble learning techniques and semi-supervised learning also showed promising results in improving detection accuracy and reducing the rate of positive errors. The study emphasizes the importance of adaptive and sustainable approaches in the development of machine

ARTICLEINFO

Article History:
Submitted:
Received:
Accepted:

Keywords:
Cybersecurity, Cloud Computing, Machine Learning, Algorithm Optimization, Literature Studies.

learning-based security solutions tailored to the dynamic characteristics of cloud computing environments. These findings are expected to make a theoretical and practical contribution to the development of a more resilient and intelligent cybersecurity system in the digital era.

1. INTRODUCTION

The development of cloud computing technology has made a significant contribution to efficiency and flexibility in managing data and IT services in various sectors, ranging from government, business, to education (Mell & Grance, 2011). However, as the adoption of cloud computing increases, the risk of cyberattacks is also increasing, especially in terms of data security, user privacy, and system integrity (Zhou et al., 2010). Threats such as Distributed Denial of Service (DDoS), phishing, malware, and unauthorized access are now crucial issues that require a new approach to security management (Hashizume et al., 2013).

Previous research has extensively examined the application of machine learning algorithms in detecting and counteracting cyberattacks (Buczak & Guven, 2016; Sahoo et al., 2017). However, most studies still focus on applying algorithms directly without optimizing the parameters and model architectures used, so their effectiveness in the context of cloud computing is still limited (Nguyen et al., 2020). This is an important research gap to bridge.

The urgency of this research is based on the fact that the cloud computing environment is dynamic and multitenancy, so conventional security systems become less adaptive and responsive (Modi et al., 2013). Therefore, a smarter and adaptive approach is needed through the optimization of machine learning algorithms to improve threat detection capabilities in real-time (Chiba et al., 2016).

Although there are various algorithms such as Support Vector Machine (SVM), Random Forest, and Deep Neural Networks that have been used in intrusion detection systems (Revathi & Malathi, 2015; Kim et al., 2016), this study offers a novelty in the form of a systematic approach to optimize these algorithms based on threat characteristics in the cloud environment, more representative use of data, and integrated tuning and ensemble learning techniques.

The purpose of this study is to examine and critically analyze various machine learning algorithm optimization strategies in improving cybersecurity in cloud computing environments. Through a literature study approach, this research is expected to provide theoretical contributions in the form of conceptual frameworks as well as practical contributions as a reference for the development of a more effective and efficient cloud-based cybersecurity system.

The benefits of this research include providing new insights for researchers and practitioners in the field of cybersecurity, enriching academic literature related to the optimization of machine learning algorithms in the context of cloud computing, and becoming the basis for the development of smart technology-based security policies and systems.

Definition of Machine Learning Algorithms

Machine learning algorithms are a set of computational methods that allow systems to learn from data and make decisions or predictions without the need for explicit programming. In other words, this algorithm works by recognizing patterns in the available data, then using

it to make decisions or complete certain tasks automatically (Mitchell, 1997). This learning process involves training the model using a specific dataset, then testing the model's ability to predict or classify new data. In the context of modern information technology, machine learning is an important component in the development of intelligent systems.

Types of Machine Learning Algorithms

In general, machine learning algorithms are divided into three main categories, namely supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, the model is trained using labelled datasets, such as email spam classification or network intrusion detection, using algorithms such as Support Vector Machine (SVM), Random Forest, or Logistic Regression. Unsupervised learning is used when data has no labels, and algorithms such as K-Means Clustering or Principal Component Analysis (PCA) are used to find hidden patterns or structures in the data. Meanwhile, reinforcement learning involves agents learning from interactions with the environment through feedback systems in the form of rewards and punishments, and is commonly used in autonomous systems such as robotics or dynamic decision-making.

The Role of Machine Learning Algorithms in Cybersecurity

In the context of cybersecurity, machine learning algorithms are used to detect anomalies, recognize attack patterns, and respond to threats automatically. For example, by training the model using network activity log data, the system can detect suspicious activity that may be a DDoS attack, malware, or an illegal access attempt. Algorithms such as Deep Neural Networks and ensemble learning have proven to be effective in improving detection accuracy and reducing false positives. The advantage of machine learning in security lies in its ability to continuously adapt to new types of attacks through the updating of training data, making it an adaptive and relevant solution in the face of the ever-evolving cyber threat landscape (Sommer & Paxson, 2010).

2. METHODS

This research uses a qualitative approach with the type of literature review or library research, which aims to explore, review, and analyze various previous studies related to the optimization of machine learning algorithms in improving cybersecurity in the cloud computing environment. Literature studies were chosen because they were able to provide in-depth conceptual understanding and identify patterns of scientific thought from various relevant sources (Zed, 2008). This research is exploratory and analytical, with a focus on identifying research gaps, developing conceptual frameworks, and compiling synthesis of previous findings that can be used as a basis for the development of adaptive cybersecurity systems.

The data sources in this study consist of secondary data obtained from various scientific publications such as reputable international journal articles, conference proceedings, academic books, technical reports, and online repositories such as IEEE Xplore, SpringerLink, ScienceDirect, and Google Scholar. Inclusion criteria for resource selection include: (1) relevance to the topics of machine learning, cybersecurity, and cloud computing; (2) published in the last five to ten years; and (3) have academic validity verified through a peer review system.

The data collection technique is carried out through systematic searches using specific keywords such as "machine learning optimization," "cloud security," and "cyber threat detection," which are combined with snowballing techniques to search for references from the main articles. The collected data is then analyzed using the content analysis method, which is by identifying the main themes, compiling thematic categories, and comparing approaches, findings, and conclusions from each source (Krippendorff, 2018). This process is carried out inductively, namely by building an understanding based on the interpretation of the available information, and deductively by connecting the results of the analysis to the relevant theoretical framework. Thus, this research is expected to be able to provide a comprehensive synthesis of knowledge about machine learning algorithm optimization strategies in improving cybersecurity in the cloud computing environment.

3. RESULTS AND DISCUSSION

The following literature data is the result of a selection of a number of literature found through systematic searches on scientific databases such as IEEE Xplore, SpringerLink, ScienceDirect, and Google Scholar. From a total of dozens of articles found, screening was carried out based on the inclusion criteria that have been set, namely: (1) the relevance of the topic to the optimization of machine learning algorithms in the context of cybersecurity in the cloud computing environment, (2) published between 2015–2024, and (3) coming from sources that have gone through a peer-review process. The final results of this selection process produced 10 core articles that were used as the basis for thematic analysis in this study. The following table presents a summary of the literature data, including the author's name, year, research title, algorithm used, optimization method, and key findings.

Table 1. Summary of Selected Literature Data

Author & Year	Article Title	Key findings
Buczak & Guven (2016)	A Survey of Data Mining and Machine Learning Methods for Cyber Security	*Intrusion detection is improved with feature optimization.*
Nguyen et al. (2020)	Improving IDS Performance through Hybrid Feature Selection and Ensemble	Accuracy increases, false positives decrease
Chiba et al. (2016)	Intelligent Approach to Build Sustainable Cloud IDS	Effective on large-scale cloud data.
Sahoo et al. (2017)	A Survey on Intrusion Detection in Cloud Environment	Performance is better than a single approach.
Kim et al. (2016)	A Hybrid IDS Integrating Anomaly and Misuse Detection	*Faster and more accurate detection.*

Revathi & Malathi (2015)	Analysis on NSL-KDD Using Various ML Techniques	*Model validation is important for detection accuracy.*
Alrawashdeh & Purdy (2016)	Deep Learning for Intrusion Detection in Cloud	*Efficiently detect zero-day attacks.*
Vinayakumar et al. (2019)	Deep Learning Approaches for Cybersecurity	*Architecture in influencing system performance.*
Ambusaidi et al. (2016)	Feature Selection for Intrusion Detection Using Gain Ratio and Chi-square	*The selection of features increases the efficiency of the classification process.*
Diro & Chilamkurti (2018)	DL-based Intrusion Detection for Cloud Security	*High accuracy with optimal training time.*

From the table presented, it can be seen that there is a strong tendency towards the use of Supervised Learning algorithms such as Support Vector Machine (SVM), Random Forest, and Naive Bayes in intrusion detection systems (IDS) based on cloud computing. This suggests that these algorithms are still the top choice due to their ability to produce accurate classifications when trained with labeled data. Some studies have even combined these algorithms in an ensemble approach to improve accuracy and minimize classification errors, as Nguyen et al. (2020) and Sahoo et al. (2017) have done.

In addition to conventional algorithms, recent studies have shown an increase in the use of Deep Learning algorithms, such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Autoencoder. Studies by Alrawashdeh & Purdy (2016) and Diro & Chilamkurti (2018) show that Deep Learning can detect zero-day attacks more efficiently due to its ability to capture complex patterns in big data. This approach is also more adaptive to the changing cyber threat environment, which is particularly relevant in the context of dynamic cloud computing.

The optimization techniques used in various studies are quite diverse, including feature selection, ensemble learning, parameter tuning, and network architecture optimization. Methods such as Gain Ratio and Chi-Square used by Ambusaidi et al. (2016) have proven to be effective in filtering out the most relevant features for attack detection, thereby reducing the computational load while improving system efficiency. On the other hand, Vinayakumar et al. (2019) emphasize the importance of network architecture design in Deep Learning so that the system is not only accurate, but also stable and fast in the training process.

The hybrid approach, which is combining two or more techniques in a single system, is widely used to maximize the advantages of each method. A study by Kim et al. (2016) shows that the combination of anomaly-based detection and signature-based detection provides more effective results in detecting known and new threats. This strategy is particularly relevant in a multi-user cloud environment and is vulnerable to different types of attacks.

Almost all studies state that machine learning algorithm optimization can significantly improve attack detection accuracy, reduce false positive rates, and speed up system response times. However, the effectiveness of the system also depends heavily on the quality of the training data and the selection of the right features. A study by Revathi & Malathi (2015)

emphasizes the importance of model validation using standard datasets such as NSL-KDD so that the results can be objectively measured and replicated.

From this interpretation, it can be concluded that machine learning algorithm optimization not only provides improved technical performance, but also offers a smarter and more adaptive approach in the development of cloud-based cybersecurity systems. This approach is able to answer the challenges of cloud environments that are dynamic, complex, and full of uncertainty. Therefore, the synthesis of this literature provides a strong conceptual foundation for the development of a cloud-based IDS system that is not only accurate and efficient, but also scalable and sustainable.

Discussion

In the current era of digital transformation, the use of cloud computing services continues to increase significantly, both in the business, education, and government sectors. However, this development has also been followed by an increase in increasingly complex cybersecurity threats, including DDoS attacks, phishing, malware, and data breaches that exploit gaps in cloud infrastructure. Therefore, a new, more adaptive, automated, and intelligent approach is needed to detect and respond to these threats in real-time.

Machine learning (ML) is a very potential solution in answering these challenges because it is able to study threat patterns from historical and actual data, and is able to generalize to attack conditions that have never been recognized before. ML also has advantages in attack detection automation, so it can reduce reliance on traditional reactive and rigid security systems. In line with the theory of Artificial Intelligence-based Security Systems, AI/ML-based systems can proactively and dynamically improve cyber resilience (Buczak & Guven, 2016).

From the results of the literature study, it can be seen that algorithms such as Support Vector Machine (SVM), Random Forest, and Naive Bayes dominate the application in intrusion detection. This algorithm is proven to perform well in labeled data classification and is capable of processing large amounts of data with relatively efficient computational time. The authors argue that the selection of this algorithm is also based on its technological maturity and ease of application in the cloud-based cybersecurity system environment.

The findings also show that Deep Learning approaches, such as CNN, RNN, and Autoencoder, are starting to be widely used in recent research. This indicates a shift towards a more complex and adaptive approach to handling non-linear and dynamic data patterns. Deep learning is particularly relevant to apply to cloud architectures due to its scalability and ability to handle large and varied amounts of data. A study by Alrawashdeh & Purdy (2016) reinforces this by proving its ability to detect zero-day attacks that are very difficult to detect by conventional systems.

Commonly used optimization techniques such as feature selection and ensemble learning have been proven to improve the efficiency and accuracy of the detection system. Feature selection helps eliminate irrelevant features so that the classification process becomes lighter and faster. While ensemble learning combines the power of multiple algorithms to balance between bias and variance in the model. The authors view that the combination of these two techniques is an ideal solution to improve the performance of cybersecurity systems without having to significantly increase the computing load.

Some studies propose a hybrid approach, which combines anomaly-based detection and signatures. This model has proven effective in detecting both familiar and new attacks. In the cloud context, a hybrid approach is critical because cloud systems are multi-tenant and dynamic, so a single method is often not enough to handle the complexity of emerging attacks. The author's opinion is in line with this approach, because practically cybersecurity must be flexible and adaptive to the ever-changing threat dynamics.

These findings are also in line with cloud security standards such as ISO/IEC 27017 and the principles in Zero Trust Architecture (ZTA), which emphasize the importance of behavior-based and data-driven security controls. Machine learning provides the technological basis for ZTA implementation because it is able to model user and device behavior to determine the level of risk contextually. Thus, ML not only detects attacks, but also serves as the foundation for more modern risk-based security policies.

Although many algorithms show high performance in experimental environments, the main challenge in real-world implementation is the quality and diversity of training data. Unrepresentative or unbalanced data can create model bias and increase false positive rates. The authors consider that the development of an optimal intrusion detection system must also consider a mature data collection and validation strategy so that the prediction results are reliable.

These findings provide important implications for information security practitioners in cloud computing environments. Machine learning algorithm optimization is not only a technical issue, but also a long-term strategy in building a responsive, automated, and sustainable cybersecurity system. The implementation of ML-based detection systems can reduce the burden on security teams, improve operational efficiency, and significantly lower the risk of losses from successful cyberattacks.

As a reflection, the authors see that the merger between the power of machine learning algorithms and cloud-native architecture approaches will become the dominant trend in cybersecurity systems in the future. However, challenges such as the need for model interpretability (explainable AI) and the protection of training data privacy should also receive further attention in future research. Thus, the results of this literature study not only provide an overview of the current situation, but also open up space for exploration for the development of smarter and more humane cloud security systems.

4. CONCLUSION

This study shows that machine learning algorithm optimization is a very potential approach in improving cybersecurity systems in cloud computing environments. Algorithms such as Support Vector Machine (SVM), Random Forest, Naive Bayes, and deep learning-based approaches such as CNN and RNN have been proven to be able to detect various types of attacks, including zero-day attacks, with high accuracy. This advantage is crucial given the characteristics of a dynamic, open, and vulnerable cloud to complex and massive attacks.

In addition to algorithm selection, the application of optimization strategies such as feature selection, ensemble learning, and hybrid detection models has been proven to improve system efficiency and reduce the rate of error detection. The results of the literature review show that the combination of these techniques can create a more intelligent, adaptive, and effective intrusion detection system in cloud environments that have a high data traffic load.

These findings confirm that machine learning-based approaches not only provide added value technically, but are also strategically relevant in cloud-based cyber risk management.

However, the application of machine learning algorithms still faces challenges, especially in terms of the availability of quality and representative training data, as well as the need for models that can be interpreted transparently by users or security administrators. Therefore, the development of models that are explainable, efficient in computing, and can be integrated with existing cloud security frameworks need to be the focus of further research.

Further Research Recommendations

Future research is suggested to develop a machine learning-based intrusion detection system that is real-time, explainable, and cloud-native. The main focus should be on combining machine learning techniques with policy-based security approaches such as Zero Trust Architecture. In addition, testing the model on real data from various public and hybrid cloud platforms will strengthen the external validity of the system being built. Researchers also need to pay attention to ethical and data protection aspects, especially in the context of processing sensitive data in a multi-tenant cloud environment. With these steps, the contribution of machine learning in cloud security can be more optimal, reliable, and sustainable.

5. ACKNOWLEDGMENT

The author would like to thank all parties who have provided support in the preparation of this article. Special awards are presented to academic institutions and digital libraries that have provided access to various relevant literature sources, which are the main basis for the implementation of this literature study. Thanks are also extended to previous researchers whose work has become an important reference in building the framework of thought and analysis in this article. Without the scientific contribution of the research community that has studied the topic of machine learning and cloud computing security, this study would not be able to be compiled in depth and comprehensively.

6. AUTHORS' NOTE

This article was compiled as part of the author's commitment to contribute to the development of scientific literature in the field of information technology, especially on the application of machine learning algorithms to improve cybersecurity in cloud computing environments. All articles are the result of systematic literature reviews, while upholding academic principles and scientific integrity. The author has no conflicts of interest related to this topic, and does not receive any external funding in the writing process of this article. The author is open to criticism, input, and academic collaboration from other researchers to expand the scope and impact of future studies.

7. REFERENCES

Buczak, A. L., & Guven, E. (2016). A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. IEEE Communications Surveys & Tutorials, 18(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502

Chiba, Z., Abghour, N., Moussaid, K., & Rida, M. (2016). Intelligent Approach to Build a Sustainable Cloud Intrusion Detection System. Procedia Computer Science, 83, 1164–1169. https://doi.org/10.1016/j.procs.2016.04.245

Hashizume, K., Rosado, D. G., Fernández-Medina, E., & Fernandez, E. B. (2013). An Analysis of Security Issues for Cloud Computing. Journal of Internet Services and Applications, 4(1), 5. https://doi.org/10.1186/1869-0238-4-5

Kim, G., Lee, S., & Kim, S. (2016). A Novel Hybrid Intrusion Detection Method Integrating Anomaly Detection with Misuse Detection. Expert Systems with Applications, 41(4), 1690–1700. https://doi.org/10.1016/j.eswa.2013.08.066

Krippendorff, K. (2018). Content Analysis: An Introduction to Its Methodology (4th ed.). SAGE Publications.

Mell, P., & Grance, T. (2011). The NIST Definition of Cloud Computing. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-145

Modi, C., Patel, D., Borisaniya, B., Patel, A., & Rajarajan, M. (2013). A Survey of Intrusion Detection Techniques in Cloud. Journal of Network and Computer Applications, 36(1), 42–57. https://doi.org/10.1016/j.jnca.2012.05.003

Nguyen, T. T., Nguyen, T. D., Nguyen, D. T., & Huynh, V. N. (2020). Improving Intrusion Detection System Performance through Hybrid Feature Selection and Ensemble Learning. Applied Intelligence, 50, 3077–3093. https://doi.org/10.1007/s10489-020-01680-5

Revathi, S., & Malathi, A. (2015). A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection. International Journal of Engineering Research & Technology, 4(6), 184–187.

Sahoo, S. R., Mohapatra, D. P., & Lath, R. (2017). A Survey on Intrusion Detection in Cloud Environment. Journal of Network and Computer Applications, 77, 18–35. https://doi.org/10.1016/j.jnca.2016.10.005

Zed, M. (2008). Literature Research Methods. Jakarta: Yayasan Obor Indonesia.

Zhou, M., Zhang, R., Xie, W., Qian, W., & Zhou, A. (2010). Security and Privacy in Cloud Computing: A Survey. 2010 Sixth International Conference on Semantics, Knowledge and Grids, 105–112. https://doi.org/10.1109/SKG.2010.43