Exploring the impact of AI-driven solutions on cybersecurity adoption in small and medium enterprises

Geeta Sandeep Nadella *, Hari Gonaygunta, Deepak Kumar and Priyanka Pramod Pawar

Department of Information Technology, University of the Cumberlands, Williamsburg, Kentucky, USA.

World Journal of Advanced Research and Reviews, 2024, 22(01), 1190–1197

Publication history: Received on 07 March 2024; revised on 20 April 2024; accepted on 22 April 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.22.1.1185

Abstract

This study delves into the intricate dynamics of cybersecurity adoption within small and medium enterprises (SMEs), with a specific focus on the transformative potential of AI-driven solutions from a machine learning perspective. Leveraging a rich and diverse dataset of real-world cybersecurity incidents tailored to the SME context, the research meticulously investigates the efficacy of three prominent machine learning models: logistic regression, random forest, and gradient boost classifiers. Through a comprehensive evaluation using a suite of performance metrics, including accuracy, precision, recall, and F1 scores, the study scrutinizes the predictive capabilities of these models in forecasting and categorizing cyber threats encountered by SMEs. The findings unveil nuanced insights into the multifaceted challenges and opportunities inherent in the cybersecurity landscape of SMEs, shedding light on the complex interplay between AI-driven solutions and evolving cyber threats. While the analysis reveals commendable performance across the models, it also uncovers inherent limitations in accurately discerning and categorizing specific types of attacks. These findings underscore the critical need for ongoing refinement and optimization of AI-driven cybersecurity solutions, necessitating a continuous, iterative process informed by real-world data and adaptive learning mechanisms. By harnessing the strengths of machine learning and embracing a proactive stance towards cybersecurity, SMEs can fortify their defense mechanisms and safeguard their digital assets in the increasingly volatile and interconnected digital ecosystem.

Keywords: Artificial Intelligence; Cyber Security; Machine Learning; Random Forest; Logistic Regression; Gradient Boost; Small and Medium-Sized Enterprises (SMEs)

1. Introduction

The rapid proliferation of digital technologies and the ubiquity of internet-connected devices have ushered in an era of unprecedented opportunities and vulnerabilities in the realm of cybersecurity [1]. Cybersecurity, a critical field of study, safeguards organizations, interactions, systems, and information from a myriad of cyber-attacks [2]. The integration of the Internet of Things (IoT) holds immense promise for innovative societal applications. However, it also heightens the need for robust cybersecurity measures as financial institutions and other critical sectors face increasing cyber risks [3]. In response to these evolving threats, researchers have proposed AI-driven cybersecurity approaches that leverage advanced techniques, such as the K-NN algorithm combined with enhanced encryption and decryption procedures, to mitigate network attacks that employ information poisoning and computational intelligence [4]. The incorporation of artificial intelligence (AI) into cybersecurity solutions has emerged as a pivotal development, transforming the landscape of cyber defense.

Historically, machine learning has been utilized in cybersecurity since the 1990s, but advancements in data and computing have propelled AI to become a crucial component of modern cyberdefense strategies [5], [5]. Automated cyber defenses can now mimic human intelligence and behavior, enabling faster detection of network vulnerabilities...
compared to manual efforts [6]. The COVID-19 pandemic has further accelerated the adoption of technology, leading to a rapid increase in cybercrime and putting businesses and individuals at risk [7]. Small and medium enterprises (SMEs) are particularly vulnerable to the proliferation of cybercrime, as they often lack the resources and maturity to implement robust cybersecurity measures [8]. According to recent estimates, the global cost of cybercrime is projected to reach $10.5 trillion by 2025, highlighting the urgency for SMEs to address this growing threat [9]. Centralized cybersecurity management and the availability of adequate IT resources have been shown to enhance an organization’s security posture, underscoring the need for SMEs to adopt a more strategic and proactive approach to cybersecurity [10].

The evolving cybersecurity landscape and the increasing prominence of AI-driven solutions have prompted researchers to investigate the factors that influence cybersecurity decision-making and adoption within SMEs. The extended Technology-Organization-Environment (TOE) framework proposed by Wallace et al. [11] provides a multidimensional perspective on the determinants of cybersecurity adoption, including new variables, practice standards, cyber catalysts, and additional dimensions. Similarly, Hasan et al. [12] explored the relationship between cybersecurity readiness characteristics and organizational performance using the TOE framework, highlighting the importance of a comprehensive approach to cybersecurity management. This study aims to contribute to the growing body of knowledge by exploring the impact of AI-driven solutions on cybersecurity adoption within small and medium enterprises, leveraging a machine learning-based approach to uncover insights and inform future research and practical applications.

2. Literature Review

Many cybercrimes, including hacking attacks and data breaches, have received extensive media coverage, elevating the subject of cybersecurity to the forefront of public discourse. According to Zimmerman and Renaud’s (2019) investigation, rather than depending on technological means, they concentrated on improving aspects that lead to resilient cybersecurity and positive outcomes [13]. According to Liu et al. (2020), a company’s cybersecurity posture improves when management is centralized [14]. Benz and Chatterjee (2020) noted that small and medium enterprises (SMEs) are among the most fragile and immature business models, and the ever-increasing cyber hazards for SMEs are a direct result of the proliferation of cybercrime, as elucidated [15], [16].

The availability and utilization of an organization’s IT infrastructure are critical to the smooth running of its internal business processes. Hasan et al. (2021) noted that when organizations assess the availability and utilization of IT, they must consider the investment, capacity, and hardware components [17]. According to three studies, having enough IT resources makes systems more secure and reduces the likelihood of security breaches and incidents [18]. Decisions regarding the implementation of cybersecurity are significantly impacted by the revised and extended TOE framework that was presented by Wallace et al. in 2020. This framework includes new variables, practice standards, cyber catalysts, and extra dimensions under the classic organization, technology, and environment dimensions [19]. Hasan et al. (2021) investigated the relationship between different cybersecurity readiness characteristics and organizational financial and non-financial performance by using the TOE framework [17]. Brunner et al. (2021) assessed the state of risk management techniques in connection with Information Security Management Systems (ISMS) for regions of Austria, Germany, and Switzerland, examining how companies maintain data security by using standard operating procedures, data-gathering methods, tool types, interpersonal arrangements, and documentation artifacts [20]. Cyber-physical systems, also known as the CPS Environment, aim to create IoT infrastructure for educational research purposes in a variety of PeS-related domains, with the primary objective of providing physical architecture and facilitating the exploration of its possible practical uses by scholars and learners [21].

Studies have examined how investors perceive a company’s likelihood of declaring bankruptcy in the future in relation to eXtensible Business Reporting Language (XBRL). Standards for implementing XBRL reduce the expected risk of crashes and companies with more volatile performance, opaque accounting, and untrustworthy projections from analysts [22]. These problems likewise impact procedures used in Cyber Digit manufacturing (a DM), and creating a meta-architecture might help resolve this problem [23]. A monetary risk related to cybersecurity is highlighted in a hypothetical situation using the College of Industry and Computer Engineering’s Explorers application, where the evolutionary algorithm, or GA, is used to pick the components of the protective mechanism, and a fuzz assessment is used to determine the degree of fitness [24].
3. Methodology

The methodology employed in this research explores the impact of AI-driven solutions on cybersecurity adoption within small and medium enterprises (SMEs) from a machine learning perspective. It begins with the collection of a dataset sourced from cybersecurity attacks, tailored to SME contexts, followed by preprocessing steps to handle missing values and ensure data quality. Descriptive statistics are computed to analyze numerical attributes, while exploratory data analysis techniques uncover patterns and trends in the dataset, focusing on indicators of compromise and response actions taken [25]. A temporal analysis is conducted to understand the dynamics of cyber threats over time. Predictive models that leverage machine learning are developed to measure the efficacy of AI-driven solutions in mitigating cyber-attacks.

In the machine learning modeling phase, three distinct algorithms - logistic regression, random forest, and gradient boost - are employed to analyze the dataset and predict cyber threats within SMEs [26]. Logistic regression, a classic statistical method, is utilized to model the probability of occurrence of a binary outcome, making it suitable for classification tasks such as identifying malicious activities. Meanwhile, random forest, a powerful ensemble learning technique, utilizes a multitude of decision trees to generate robust predictions by aggregating outputs of multiple individual models [28]. Gradient boost, a cutting-edge boosting algorithm, revolutionizes cybersecurity modeling for SMEs by iteratively refining predictive performance through sequential optimization [27].

![Proposed Framework](image)

**Figure 1** Proposed Framework

The methodology employed in this research explores the impact of AI-driven solutions on cybersecurity adoption within small /medium enterprises (SMEs) from a machine learning perspective. It begins with the collection of a dataset sourced from cybersecurity attacks, tailored to SME contexts, followed by preprocessing steps to handle missing values and ensure data quality [28]. Descriptive statistics are computed to analyze numerical attributes, while exploratory data analysis techniques uncover patterns and trends in the dataset, focusing on indicators of compromise and response actions taken. A temporal analysis is conducted to understand the dynamics of cyber threats over time. Predictive models that leverage machine learning are developed to measure the efficacy of AI-driven solutions in mitigating cyber-attacks. Model performance is evaluated using appropriate metrics, and findings are interpreted to provide insights into enhancing cybersecurity resilience in SME environments, accompanied by recommendations for future research and practical implications [28].

3.1. Data Collection

In this research, data collection is paramount for exploring the impact of AI solutions on cybersecurity adoption in enterprises of medium and small buildings. The dataset utilized is carefully curated from cybersecurity attacks, explicitly focusing on incidents pertinent to SMEs. This dataset comprises various attributes crucial to understanding cyber incidents, like the time stamp, protocol, packet length, packet type, source-destination IP addresses, source-port
numbers for the destination, and Traffic Type, among others, all tailored to SME cybersecurity contexts. By incorporating these attributes, the dataset enables a comprehensive analysis of cybersecurity threats faced by SMEs, providing insights into the nature, frequency, and severity of cyberattacks targeting SME businesses. Through this data collection process, the research lays the foundation for examining the efficacy of AI-driven solutions in bolstering cybersecurity defenses within the SME sector, offering valuable insights to inform future cybersecurity strategies and practices.

![Figure 2 Cybersecurity Dataset](image)

### 3.2. Data Preprocess

Information preprocessing is a vital step in preparing datasets for analysis, particularly in the context of exploring impact AI-driven solutions on cyber-security adoption in small-medium enterprises from a machine learning perspective [29]. Initially, the dataset undergoes preprocessing to address missing values and ensure completeness of data. Columns such as Timestamp are converted to date time format to facilitate temporal analysis and enable exploration of trends over time. Moreover, to convert categorical variables in a set of data into numbers that are appropriate for algorithmic machine learning, methods of encoding, such as labeling and single-hot encoding, can be applied. The algorithms for learning ML guarantee that it will work by allowing for the inclusion of categorical attributes in analysis. Additionally, outlier detection techniques may be applied to identify and handle anomalies within data, thereby enhancing data quality and robustness of subsequent analyses. Via accomplishment preprocessing stages, the dataset is refined and optimized for more examination and modeling, facilitating a comprehensive investigation into AI-driven solutions on cyber-security adoption in SMEs [30].

### 4. Data Analysis of Malicious Activity

In an analysis of malicious activity within small and medium enterprises (SMEs), the dataset is scrutinized to identify patterns and characteristics indicative of cyber threats. This examination primarily focuses on various indicators of compromise, such as IDS/IPS Alerts, Malware Indicators, and Alerts/Warnings. By delving into the indicators, the research aims to understand the nature and severity of threats faced by SMEs in their cybersecurity landscape. Additionally, attention is given to discerning top users and network segments associated with these malicious activities, providing insights into potential vulnerabilities within SME networks. Through this analysis, a comprehensive understanding of the threat landscape within SMEs is obtained, facilitating the development of effective cybersecurity strategies and the adoption of AI-driven solutions to mitigate and counteract malicious activities.
In Figure 3, the distribution of attack types within small and medium enterprises (SMEs) reveals intriguing insights into prevailing cybersecurity threats. Among recorded attack types, disseminated denial attacks constitute a significant portion, accounting for 12,500 instances. DDoS attacks, known for their disruptive nature, pose substantial risks to SMEs by overwhelming their network infrastructure and rendering services inaccessible to legitimate users. Concurrently, malware-based attacks totaling 12,300 instances represent another substantial threat vector. A wide range of harmful software is used in virus assaults, with the goal of infiltrating networks, compromising the confidentiality of information, and interfering with corporate processes. Intrusion attempts, comprising 13,000 instances, constitute a notable portion of the attack landscape. Intrusions involve unauthorized access attempts into network systems, potentially leading to data breaches, system compromise, and unauthorized data access.

In Figure 4, the proportion of severity levels across all network segments showcases a balanced distribution, with each severity level constituting approximately one-third of the total distribution. Specifically, analysis reveals that 33% of incidents are categorized as low severity, 34% as high severity, and another 33% as medium severity. This equilibrium suggests a diverse range of cybersecurity incidents occurring across SME network segments, with varying degrees of impact and criticality. The prevalence of high-severity incidents highlights the potential for significant disruptions and damage to SME operations and assets.

4.1. Machine Learning Modeling

In the machine learning modeling phase, three distinct algorithms—logistic regression, random forest, and gradient boost—are employed to analyze datasets and predict cyber threats within small and medium enterprises (SMEs). Logistic regressions, a classic arithmetical method, is utilized to model the phase likelihood of occurrence of a binary outcome, making it suitable for classification tasks such as identifying malicious activities. Meanwhile, random forest, a powerful collaborative knowledge technique, utilizes a multitude of decision trees to generate robust predictions by aggregating outputs of multiple individual models.
Logistic Regression: In the background of cybersecurity analysis for small and medium enterprises, logistic regression serves as a foundational model for predicting and classifying cyber threats. Because of its simplicity and interpretability, logistic regression is used to estimate the probability of a binary outcome, making it suitable for identifying malicious activities within SME networks. By analyzing various features and attributes from the dataset, logistic regression computes the likelihood of specific cyber events occurring, providing valuable insights into potential vulnerabilities and threat vectors. Despite its linear nature, logistic regression remains a valuable tool in the cybersecurity arsenal, particularly for scenarios where transparency and explainability are paramount.

Random Forest: Random forest, a versatile ensemble learning technique, offers SMEs a robust approach to cybersecurity modeling by harnessing the power of multiple decision trees. Individually, the D-tree within the random-forests ensemble autonomously learns from different subsets of the dataset, capturing unique patterns and relationships between features. Through aggregation of predictions from individual trees, random forest generates highly accurate and reliable forecasts while mitigating the risk of overfitting. Logistic regression is particularly effective in handling the complex and heterogeneous data characteristic of cybersecurity incidents in SMEs. Leveraging the shared understanding of manifold decision trees with random forests, SMEs can discern subtle nuances in cyber threat landscapes and proactively fortify their defense mechanisms.

Gradient Boost: Gradient boost, a cutting-edge boosting algorithm, revolutionizes cybersecurity modeling for SMEs by iteratively refining predictive performance through sequential optimization. By constructing a sequence of weak learners—typically decision trees—gradient boost continuously focuses on minimizing prediction errors, gradually improving the model correctness with apiece repetition. This iterative learning procedure enables gradient boost to imprison multifaceted connections plus non-linear relationships within the dataset, leading to enhanced predictive capabilities. Additionally, gradient boost adapts its focus to prioritize misclassified instances, allowing SMEs to effectively allocate resources towards addressing the most critical cyber threats. Through its adaptive and data-driven approach, gradient boost empowers SMEs to stay ahead of evolving cyber threats and bolster their cybersecurity defenses with precision and efficacy.

4.2. Results Evaluation

In the results evaluation phase, the performance of logistic regression, random forest, and gradient boost models is meticulously assessed to ascertain their effectiveness in predicting and mitigating cyber threats within small and medium enterprises (SMEs). The predictive capability and generalization proficiency of these algorithms are quantitatively evaluated using a comprehensive suite of metrics, including precision, recall, recall accuracy, F1 scores, and the area under the receiver operating characteristic curve (AUC-ROC). By juxtaposing these metrics across different models, businesses can derive nuanced insights into the relative advantages and disadvantages, as well as the potential trade-offs and challenges, associated with each approach. This analytical framework facilitates informed decision-making by highlighting the most productive strategies for enhancing cybersecurity postures in the context of SMEs.

Table 1 Models Comparison results

<table>
<thead>
<tr>
<th>Models-names</th>
<th>Attack-Type</th>
<th>Accuracy</th>
<th>Precisions</th>
<th>Recall</th>
<th>F1-S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>DDoS</td>
<td>0.34</td>
<td>0.34</td>
<td>0.48</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Intrusion</td>
<td>0.33</td>
<td>0.19</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Malware</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>Random Forest Classifier</td>
<td>DDoS</td>
<td>0.33</td>
<td>0.32</td>
<td>0.36</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>Intrusion</td>
<td>0.33</td>
<td>0.31</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Malware</td>
<td>0.32</td>
<td>0.31</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>Gradient Boosting Classifier</td>
<td>DDoS</td>
<td>0.33</td>
<td>0.33</td>
<td>0.41</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>Intrusion</td>
<td>0.34</td>
<td>0.26</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Malware</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td></td>
</tr>
</tbody>
</table>

The main results of classification models (Logistic Reg, R-Forest, and Gradient-Boost) reveal similar overall performance, with slight variations in precision, recall, and F1-score across different classes. Logistics regression attains an accuracy of 34%, while the Random Forest classifier and gradient-boosting classifier exhibit slightly lower accuracies.
of 33.5% and 33%, respectively. Upon closer examination, it becomes apparent that the models struggle to classify instances accurately across all classes, as evidenced by the relatively low precision, recall, and F1 scores for each class. Specifically, the models demonstrate challenges in distinguishing between dissimilar kinds of fake dangers, such as DDoS attacks, Intrusions, and Malware infections. Despite these limitations, the models achieve a balanced performance across the classes, as indicated by the similar macro and weighted average metrics.

5. Conclusion

In conclusion, the evaluation of logistic regression, random forest classifiers, and gradient boosting are these models implemented for predicting cyber threats within small and medium enterprises (SMEs), which underscores both the potential and challenges of utilizing machine learning in cybersecurity. While all three models exhibit comparable overall accuracies, precision, recall, and F1-scores, they demonstrate limitations in accurately classifying specific types of attacks such as DDoS, Intrusions, and Malware. Despite these challenges, the models provide valuable insights into the cybersecurity landscape of SMEs and serve as foundational tools for threat detection and mitigation. Moving forward, further optimization and refinement of these models, coupled with ongoing checking and variation near developing dangers, are indispensable to ensure the efficiency of AI-driven cybersecurity solutions in safeguarding SME networks. Additionally, the findings emphasize the importance of a holistic approach to cybersecurity, integrating advanced machine learning techniques with robust security protocols and practices tailored to the unique needs and constraints of SME environments. By leveraging the strengths of machine learning and continuously improving cybersecurity strategies, SMEs can better safeguard their electronic resources and reduce the dangers associated with cyberattacks in the networked world of tomorrow.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References


