Nigerian Journal of Technology (NIJOTECH)
Vol. 44, No. 3, September, 2025, pp.461 - 470
www.nijotech.com

Print ISSN: 0331-8443
Electronic ISSN: 2467-8821
https://doi.org/10.4314/njt.v44i3.10

LEVERAGING RANDOM FOREST ALGORITHM FOR PROACTIVE DETECTION OF INFORMATION BREACHES IN THE NIGERIAN OIL AND GAS SECTOR

AUTHORS:
*C. U. Ebelogu[1], R. Prasad[2], H. I. Bisallah[3], M. Sanusi.[4], I. Musa[5]

AFFILIATIONS:
[1-5] Department of Computer Science, University of Abuja, Abuja, Nigeria.
[1]christopher@uniabuja.edu.ng,
[2]rprasad@aust.edu.ng,
[3]hashim.bisallah@uniabuja.edu.ng,
[4]sanusi.muhammad@nasrda.gov.ng,
[5]Israelmusa9@gmail.com.

***CORRESPONDING AUTHOR:**
Email: christopher@uniabuja.edu.ng.

ARTICLE HISTORY:
Received: November 19, 2024
Revised: August 28, 2025
Accepted: September 09, 2025
Published: October 15, 2025

KEYWORDS:
Information Security, Machine Learning, Cyber Threat Detection, Nigerian Oil and Gas Sector, Random Forest Algorithm.

ARTICLE INCLUDES:
Peer review

DATA AVAILABILITY:
On request from author(s)

EDITORS:

Ozoemena Anthony Ani.

FUNDING:
None

Abstract

The Nigerian oil and gas sector's dependence on information and communication technologies makes it vulnerable to various cybersecurity threats, such as data breaches that can disrupt operations and cause significant financial losses. Despite implementing several security measures like firewalls, authentication systems, and intrusion detection systems, they struggle to keep up with the rapidly changing threat landscape. However, machine learning techniques offer potential for strengthening security in the oil and gas industry. Therefore, this study collected data on cyberattacks from relevant organizations, analyzed, cleaned, preprocessed, and split the dataset into training and testing sets using an 80/20 ratio with Python tools on Google Colab. The research identified common attack vectors, which were used to develop a machine learning model aimed at detecting and mitigating cyber incidents. The initial phase of model development faced challenges due to imbalanced datasets, leading to biased results. By applying data balancing techniques like SMOTE, the model's accuracy increased from 75% to 92%. Addressing the class imbalance also enhanced other evaluation metrics like recall, precision, and F1 score, demonstrating the model's potential for more accurate threat detection and prevention. The study concludes that integrating machine learning with cybersecurity can improve proactive security measures for infrastructure in the oil and gas sector through early warning systems with minimal false positives. This highlights the importance of advancing cybersecurity practices in Nigeria's oil and gas industry by leveraging machine learning to adapt to and counteract emerging threats effectively.

1.0 INTRODUCTION

The Nigerian oil and gas sector is a critical industry that relies heavily on information and communication technologies to manage operations, assets, and production [1]. However, this reliance on technology also exposes the sector to various information security breaches and threats [2]. High-profile information security breaches and threats in the Nigerian oil and gas sector have highlighted the need for improved cybersecurity measures [3]. The study by [4] presents a detailed analysis of cyberattacks in the oil and gas sector domain from the early 1990s to

HOW TO CITE:
Ebelogu, C. U., Prasad, R., Bisallah, H. I., Sanusi, M. and Musa, I. "Leveraging Random Forest Algorithm For Proactive Detection Of Information Breaches In The Nigerian Oil And Gas Sector", *Nigerian Journal of Technology*, 2025; 44(3), pp. 461 – 470; https://doi.org/10.4314/njt.v44i3.10

2020. The study helps in understanding the attack patterns and trends over time.

The absence of proper information security measures has been linked to issues such as infrastructure disruption, loss of information to unauthorized persons, revenue sabotage, utility interruption, terrorism, and management vulnerability [5]. Information security, or InfoSec for short, is defined by [6] as an assortment of security practices and instruments used to safeguard sensitive business data against exploitation, illegal access, interruption, or destruction. Thus, enforcing information security principles in the Nigerian oil and gas sector is very important because the sector provides a significant source of revenue for the country [7].

Various kinds of information security measures have been deployed to address these risks [8]. Traditional approaches for cyberattack mitigation, such as firewalls, intrusion detection systems, antivirus software, and access control lists, are unable to keep pace with the rapidly evolving threat landscape, leaving the sector vulnerable to new and emerging threats [9]. Machine learning can enhance the sector's information security posture by enabling real-time threat detection and response [10].

The study by [11] presented a machine-learning framework combining the Random Forest classifier and Prophet model for real-time anomaly detection and maintenance forecasting in Nigeria's oil and gas pipelines. The system achieved high accuracy (93.48%) and demonstrated effective predictive maintenance capabilities, which can potentially reduce operational downtime and mitigate environmental risks. However, limitations include prediction errors, Root Mean Squared Error (RMSE): 21.48, Mean Absolute Error (MAE): 18.17, and a Mean Absolute Percentage Error (MAPE) of 14.87%, indicating some discrepancies in forecasted values.
The review by [12] explored machine learning (ML) to enhance instrumentation accuracy in the oil and gas sector. It addresses the challenges of maintaining measurement accuracy due to environmental factors, equipment aging, and human error. The study highlights the benefits of ML, including improved data accuracy, reduced maintenance costs, and increased operational efficiency. However, the paper also notes potential challenges in integrating ML systems into existing infrastructure.

The study by [13] underscored the critical need for improved cybersecurity measures and a comprehensive policy framework to protect the Nigerian oil and gas sector from cyber threats. While the study identifies these needs, it falls short by not offering specific recommendations or a detailed analysis of the existing cybersecurity landscape. A more in-depth exploration of the challenges and potential solutions could have provided actionable insights, making the study more practical for industry stakeholders. Similarly, [14] finds that cybercrime presents a significant threat to the sector, with potential outcomes such as financial losses, reputational damage, and operational disruptions. However, like (13), this study does not offer specific recommendations for mitigating these threats or provide a detailed analysis of the sector's cybersecurity landscape. The lack of concrete strategies for addressing these challenges reduces the study's usefulness for practitioners seeking to improve cybersecurity in the industry.

The study by [15] explored the application of Machine Learning (ML) and Multi-Agent Systems (MAS) in the oil and gas industry, highlighting their potential to address challenges such as data management, efficiency improvements, and fraud prevention. However, the study's lack of comprehensive empirical data and specific insights into the implementation of these technologies weakens its contribution.
The study by [16] presents a comprehensive review of the application of ML and artificial intelligence (AI) in the oil and gas industry, particularly focusing on upstream operations.

The study highlights how ML and AI techniques enhance data processing, interpretation, and decision-making, addressing the industry's challenges related to large-scale data handling. It outlines various ML/AI methods, their benefits in improving efficiency and reducing risks, and the limitations faced in real-world applications. However, the discussion remains somewhat abstract, lacking practical examples or detailed case studies that could guide practitioners in navigating these challenges.

The study by [17] contributed to the body of literature by identifying high-performance models like identifying such models like Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Network (ANN), and Random Forest (RF), for predicting crude oil production in Nigeria. While the study provides valuable insights into forecasting methods, its reliance on secondary data and the inherent limitations of performance metrics like Root Mean Squared Error (RMSE), Mean Absolute

https://doi.org/10.4314/njt.v44i3.10

Percentage Error (MAPE), and Nash-Sutcliffe Efficiency (NSE) suggest that the findings should be interpreted with caution.

The research by [18] highlighted the transformative potential of AI, particularly ML, in revolutionizing the oil and gas industry. Their study demonstrates AI's impact across various operations, especially in seismic data processing, underscoring the technology's capacity to improve efficiency and generate insights.
The study by [19] provided a comprehensive analysis of the cybersecurity landscape in the Nigerian oil and gas sector, offering specific recommendations for addressing industry challenges. Despite its thorough approach, the study lacks a detailed assessment of the feasibility of the proposed solutions and does not delve into the potential limitations of the cybersecurity framework it suggests.

The study by [20] argues that AI-based solutions could significantly enhance the sector's cybersecurity posture by quickly detecting and mitigating threats. However, the study does not offer a detailed analysis of the current cybersecurity landscape or practical recommendations for implementing these AI solutions, limiting its contribution to the field.
[21] examined the deployment of Industry 4.0 technologies in the oil and gas upstream sector, identifying both the advantages and challenges of adoption. While the study contributes to understanding the current state of these technologies, it could benefit from a more critical analysis of the challenges and a discussion of strategies to overcome them.

Although there have been studies on information security in the Nigerian oil and gas sector, there is still a significant gap in applying machine learning within the Nigerian oil and gas sector to address security breaches. Also, the literature review reveals that certain existing systems utilized for mitigating cyberattacks exhibit high rates of false positives, leading to the generation of false alarms. Such false alarms result in operational downtime. Therefore, this study aims to bridge these gaps by developing an ML model to mitigate these threats with minimal false alarms (false positives), using known attack vectors in the industry

2.0 MATERIALS AND METHODS

a. Data

To capture the dynamics of information security breaches within the Nigerian O & G sector, this study explores relevant government agencies responsible for regulating the industry, such as the Nigerian Upstream Regulatory Commission (NURC), Nigerian Downstream and Midstream Petroleum Regulatory Authority (NMDPRA), and Nigerian National Petroleum Corporation (NNPC), in the data collection process. It also focused on the five IOCs operating in Nigeria, which include Shell Producing Development Company (SPDC), TotalEnergies, Chevron, ExxonMobil, and Eni, which are collectively responsible for producing 45 percent of oil and 40 percent of the gas in Nigeria. Integrating these agencies and companies, the study provides a comprehensive dataset that covers various aspects of breaches and threats in the Nigerian O & G sector. An overview of the dataset, showing all the features, is presented in Table 1.

Table 1: Summary of Dataset Features

Column Name	Non-Null	Data
Status	845,021	object
Access	845,021	object
Contractor Access	845,021	object
Incident Type	845,021	object
Severity	845,021	object
Timestamp	845,021	object
Device Type	845,021	object
Device Name	845,021	object
Source IP	845,021	object
Destination IP	845,021	object
Port	845,021	int64
Packet Size	845,021	int64
Duration	845,021	object
Device Location	845,021	object
Payload Content	845,021	object
Type	845,021	object
Protocol	845,021	object

b. Data Preprocessing

To make the dataset ideal for the model development and analysis, a thorough process of data cleaning and wrangling, including Transform, Outlier detection, Removal of duplicates, Ordering, and Reshaping (TOROR) techniques, has been employed to select relevant features [22]
Two types of feature engineering are carried out; the first is the timestamp column, which contains the date and time of the attack. Although pandas require this column as an object column, it is converted using the pandas datetime method, converting the column to datetime automatically changes the datatype to

datetime64. For the other categorical data, label encoding is used to assign numerical values to all instances in the categorical data. Since none of the categorical features signifies order, nominal encoding is done utilizing LabelEncoder from sci-kit-learn to transform categorical features into numerical representations [24], facilitating the integration of non-numeric data into machine learning models.
In order to achieve a good and scalable model, only features relevant to the target class "status" are selected. Although there are different types of attack vectors in the dataset, the study is limited to identifying the status of an attack, whether it was successful or failed. The distribution of the status of the attacks in the dataset is presented in Figure 1.

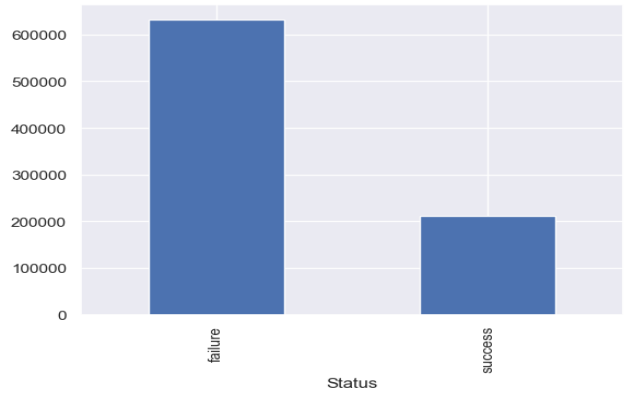

Figure 1: Distribution of the target class

The feature selection started with a correlation test, which was conducted to check for multicollinearity in the features [25]. One of the highly correlated features has to be dropped for the other, but to determine which feature to drop. The feature's importance is calculated using the mutual_info_classif method of Sklearn feature selection [26]. This reduces the number of features from 17 to 10. Before the model development, the dataset is divided into a 70% training set and 30% set to evaluate the models' performance.

c. Model

The algorithm selected for this study is Random Forest, which is identified from the literature as one of the most used and effective algorithms in predicting cybersecurity breaches [27]. This is due to the random forest's versatility in handling a mix of categorical and numerical variables. The random forest algorithm is used in this study for a classification task, it is used to classify the status of an attack, whether it is successful or not.

The key attribute of the random forest, it is an ensemble learning technique that builds multiple decision trees and merges their results to produce a prediction. The algorithm uses the concept of **bagging** (Bootstrap Aggregating), which involves:

- ✓ Selecting random forest subsets of the dataset with replacements (bootstrapping).
- ✓ Training each decision tree on a different subset.
- ✓ Aggregating the results (either through majority voting for classification or averaging for regression.

Mathematical Formulation: Given a dataset

$$D = \{(x_1, y_1), \{(x_2, y_2) \dots ((x_n, y_n)\} \quad (1)$$

where (x_1) represents the features and (y_1) represents the target label.

1. Bootstrap Sampling:

- Randomly sample m data points from D with replacement to create a subset (D_1).
- Train a decision tree (T_1) on each (D_1).

2. Prediction Aggregation:

- For classification, the final prediction y is obtained through majority voting:

$$y = mode(T_1(x), T_2(x), \dots, T_k(x)) \quad (2)$$

- For regression, it uses averaging:

$$y = \frac{1}{k} \sum_{i=1}^{k} T_i(x) \quad (3)$$

d. Evaluation Metric

Sklearn classification report and confusion matrix methods are employed to evaluate the performance of the model. The metrics and the mathematical equation are presented as follows.

i. Accuracy: Measures the proportion of correct predictions:

$$\frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Practical Insight: A high accuracy indicates the model correctly identifies both successful and failed cyber threats, crucial for minimizing security breaches.

ii. Precision: Indicates the ratio of correctly predicted positive observations to the total predicted positives:

$$\frac{TP}{TP+FP} \quad (5)$$

Practical Insight: High precision implies fewer false positives, reducing unnecessary alerts that may lead to operational disruptions.

https://doi.org/10.4314/njt.v44i3.10

iii. Recall (Sensitivity): Reflect the model's ability to identify actual positive cases:

$$\frac{TP}{TP+FN} \quad (6)$$

Practical Insight: In cybersecurity, a high recall is critical for detecting all potential threats and ensuring comprehensive protection.

iv. F1 Score: Harmonic mean of precision and recall, providing a balance between the two:

$$F1 = 2 * \frac{Precision\ x\ Recall}{Precision+Recall} \quad (7)$$

Practical Insight: A high F1 score indicates the model's robustness in maintaining a balance between identifying threats and minimizing false alarms.

v. Confusion Matrix: A table summarizing the model's performance by comparing predicted and actual classifications of positive and negative instances.

Table 2: Confusion Matrix for Binary Classification

Attacks	Predicted: Successful Attack	Predicted: Failed Attack
Actual: Successful Attack	TP	FN
Actual: Failed Attack	FP	TN

Where:
TP = True positive (correctly predicted positive instances)
FN = False negative (incorrectly predicted negative instances)
TN = True negative (correctly predicted negative instances)
FP = False positive (incorrectly predicted negative instances)

Practical Insight: Analysing the confusion matrix helps in understanding the types of errors the model is prone to (e.g., false positives vs. false negatives).

These metrics collectively provide a comprehensive evaluation of the model's performance in detecting cybersecurity breaches within the Nigerian oil and gas sector.

3.0 RESULTS

3.1 Exploratory Analysis

The analysis of the collected data indicates variation in the distribution of attack types. It is noteworthy that the high prevalence of certain types of attacks in the Nigerian oil and gas sector can be attributed to the complexity of operations, reliance on legacy systems, and extensive use of third-party vendors. A thorough understanding of the context behind these prevalent attacks will aid the sector in prioritizing security measures, enhancing employee training, and implementing advanced technologies to better protect against these threats. Figure 2 provides a detailed overview of various types of cyber incidents in the context of the Nigerian oil and gas sect

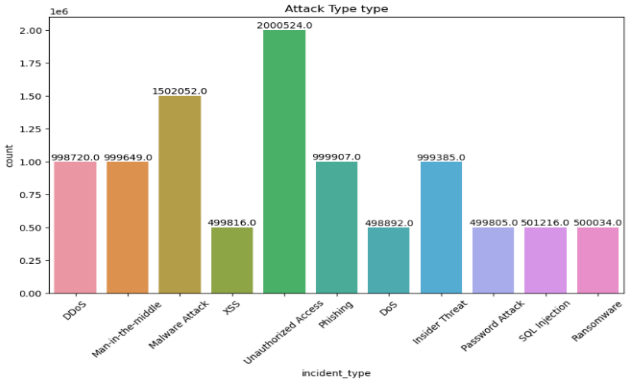

Figure 2: The Distribution of Types of Attacks

Unauthorized access is the most prevalent type of attack, with 2,000,524 incidents, which can be attributed to several factors. This could involve unauthorized individuals gaining access to sensitive operational data, proprietary technologies, or critical infrastructure controls. In the Nigerian oil and gas sector, complex and widespread infrastructure often necessitates multiple access points, making it challenging to secure all entry points effectively. Additionally, the sector's reliance on third-party vendors and contractors increases the risk of unauthorized access due to potentially lax security practices outside the primary organization's control. This type of breach can lead to significant operational disruptions, theft of sensitive data, and even physical

sabotage if critical control systems are accessed. The high frequency of these incidents suggests a pressing need for robust access control systems, such as multi-factor authentication, stringent access policies, and continuous monitoring to detect and prevent unauthorized entry. Malware attacks are the second most frequent, with 1,502,052 incidents. The high prevalence of malware in the oil and gas sector can be linked to the extensive use of legacy systems and software that may not be regularly updated or patched, creating vulnerabilities. The sector's global operations make it a lucrative target for cybercriminals looking to disrupt operations or steal proprietary information for financial gain or competitive advantage. Given the sector's reliance on interconnected systems and Supervisory Control and Data Acquisition (SCADA) networks, the deployment of advanced malware detection and prevention tools, regular system updates, and employee training on recognizing phishing attempts (a common malware delivery method) are essential. Phishing and Man-in-the-Middle (MitM) attacks are also prevalent, with 999,907 and 999,649 incidents, respectively. These attacks exploit human vulnerabilities and weak communication security. Phishing attacks are common because they are relatively easy to execute and can deceive even well-trained individuals, especially in high-pressure environments like oil and gas operations. Attackers may impersonate trusted entities to steal credentials or distribute malware. MitM attacks are prevalent due to the reliance on wireless communication and remote monitoring systems in the sector. Attackers can intercept and manipulate data between systems, leading to data breaches and operational control issues. Implementing secure communication protocols and educating employees on recognizing phishing attempts can help mitigate these risks. Insider Threats and Distributed Denial of Service (DDoS) attacks, with 999,385 and 998,720 incidents respectively, are significant concerns. Insider threats can arise from disgruntled employees or contractors who have access to critical systems and data. This necessitates strict internal controls, regular audits, and behaviour monitoring to detect and address malicious activities from within the organization. DDoS attacks, which aim to disrupt operations by overwhelming systems with traffic, can be particularly damaging in an industry dependent on continuous operations. Investing in DDoS protection solutions and creating robust incident response plans are vital. While less frequent, attacks like Denial of Service (DoS) Attacks, Password Attacks, SQL Injection, Ransomware, and Cross-Site Scripting (XSS) still pose significant risks: DoS attacks with 498,892 incidents can make essential services unavailable, impacting productivity and safety. Password Attacks with 499,805 number incidents highlight the importance of strong, unique passwords and multi-factor authentication to prevent unauthorized access. SQL Injection 501,216 incidents of attacks exploit vulnerabilities in web applications, which are often used for managing operations and data in the oil and gas sector. Ensuring secure coding practices and regular security testing can mitigate these risks. Ransomware with 500,034 incidents can cripple operations by encrypting critical data and demanding ransom for its release. The sector's reliance on data makes it a lucrative target for ransomware attacks, necessitating robust backup and recovery procedures. XSS has 499,816 incidents; it exploits web application vulnerabilities, potentially compromising sensitive information and system integrity. Adopting secure coding standards and regular vulnerability assessments is crucial to preventing XSS attacks.

3.2 Model Performance

After pre-processing the data and selecting the features, the algorithm is used to build the model. The model is built to only identify if these attacks are successful or not, not classifying each attack category as analyzed earlier. The algorithm is tested and evaluated to ensure the ideal outcome in classifying cyber threats. Table 3 shows the results of the tests done.

Table 3: Result of the Model with Unbalanced Dataset

Algorithm	Accuracy	Recall		Precision		F1 Score	
		Failure	Success	Failure	Success	Failure	Success
Random Forest	0.75	1.00	0.00	0.75	0.44	0.86	0.00

https://doi.org/10.4314/njt.v44i3.10

The result of the evaluated model presented in Table 3 shows the intrinsic nature of the dataset, as the model failed to predict successful attacks in the target class. Successful attacks constitute only 20% of all attacks. Successful attacks constitute only approximately 25.15% of the 253507 test set. The performance of the model showed it is biased towards failed attacks, which have more records than successful attacks. The model has 100% recall for the failed attacks because it recalls or remembers all failed attacks and any attack resembling them. In contrast, it has 0 recall for successful attacks, which means these models cannot remember successful attacks. However, the models also have low precision for the failed attack, with only 75% precision, and with only 44% precision for successful attacks. With these results, the confusion matrix for the model is presented in Figure 3. From Figure 3, it is shown that the classified 189484 failure instances as failures and misclassified 226 failure instances as successes. The model misclassified 63622 successful attacks as failed and only accurately classed 175 successful attacks as successful attacks due to the 44% precision.

To improve the model performance, the data distribution is balanced. The rationale for adopting this approach is based on the fact that most network attack-based datasets are unevenly distributed. The dataset is balanced using unsampling successful attacks based on the number of failed attack rows. The result of the new model is presented in Table 4.

Figure 3: The Model Predictive Behaviour

Table 4: Model Evaluation

Algorithm	Accuracy	Recall		Precision		F1 score	
		Failure	Success	Failure	Success	Failure	Success
Random Forest	0.92	0.97	0.87	0.88	0.97	0.92	0.92

The performance of the Random Forest model in this study, with an accuracy of 92% and a precision of 97% for detecting malicious network traffic, aligns with recent findings by [11], who achieved similar results using a combined RF-Prophet approach. However, while the [11] framework faced challenges with overfitting due to model complexity, the application of data balancing techniques in this study demonstrated a significant improvement in handling imbalanced datasets, thus reducing false positives to a mere 3%. This enhancement not only improves detection accuracy but also minimizes operational disruptions caused by false alarms, as emphasized by [18] in their AI-driven cybersecurity solutions.

The improved model can be deployed in real-time monitoring systems within the Nigerian oil and gas infrastructure, offering proactive threat detection capabilities. Implementing this model could lead to a reduction in unauthorized access incidents, as noted in recent studies on the effectiveness of machine learning in securing critical assets. Furthermore, integrating the model with existing SCADA systems could enhance its resilience against Distributed Denial of Service (DDoS) attacks, a prevalent threat in the sector.

Building the models with the resampled that shows different results. The models try to balance the performance metrics for each instance. As shown in Table 4, the accuracy of the random forest increased by 17%. The recall for a failed attack dropped by 3% from the previous model, the recall for a successful attack increased from 0 to 87%, the precision of the failed attack increased by 13%, while for a successful attack increased by 53%.

With these results, accurately classified 183753 failure attacks as failure attacks and misclassified only 5878 failure attacks as success; this results from the 97% recall the model has for the failure instance and a precision of 88%. The model accurately classified 165976 success attacks as success attacks and misclassified 24063 success attacks as failure attacks due to the model's low recall model's precision is higher, making it possible for the model to classify more successful instances.

https://doi.org/10.4314/njt.v44i3.10

4.0 DISCUSSION

Detecting and mitigating cyber threats has been a persistent issue for the Nigerian oil and gas industry over time, given the dynamic and ever-evolving landscape of cyberattacks. These attacks demand proactive and robust measures to curb the potentially devastating effects of their aftermath. Against this backdrop, this research study is carried out to meticulously analyze information security breaches and propose effective control model measures using advanced machine learning techniques in the Nigerian oil and gas sector, based on a comprehensive analysis of the attack dataset and the results recorded from the rigorous evaluation of the models. Through this approach, a holistic understanding of the nature, patterns, and types of attacks within the industry is gained.

Despite the promising results, the model's performance depends on data quality and distribution. Future research could explore combining ensemble learning approaches, such as boosting algorithms like XGBoost, to enhance model robustness and scalability. Additionally, integrating blockchain technology for secure data handling could further enhance data integrity in cybersecurity frameworks, providing a promising avenue for securing critical infrastructure against evolving threats.

Through thorough analysis of the data, the study identifies specific information security breaches and prevalent threats in the Nigerian oil and gas sector. It is shown that the most successful attack vector is DDoS, which exhausts the system's resources by sending overwhelming packets. The imbalanced nature of the dataset necessitated the development of two models. Initially, the models were developed with an unbalanced dataset. The performance of the algorithms with this dataset validates the "No free lunch theorem," which posits that no single algorithm is ideal for all projects; rather, its performance depends heavily on the dataset's quality and characteristics.

To improve model performance, the class weight is adjusted, and the dataset is balanced. As shown in Table 4, after balancing the dataset distribution, the performance of the three algorithms changed, once again highlighting the importance of data quality. The random forest model performs well across all evaluation metrics. However, this study does not concern itself solely with the model's generalization ability but also with how accurately individual instances in the target class are predicted. The model demonstrates high precision for detecting successful attacks and high recall for identifying failed attacks. With the model's high precision in identifying malicious activities, it achieves a low false positive rate of 3%, indicating minimal false alarms, which translates to reduced downtime and unnecessary expenses.

The results from the Random Forest model are impressive, especially when compared to those from existing literature. Some studies have shown superior accuracy, with a 92% accuracy rate in correctly classifying network traffic and a 97% chance of detecting malicious traffic. When deployed and commercialized, it is evident that this solution would provide a highly accurate and efficient model for predicting and mitigating cyber threats within the Nigerian oil and gas landscape. The ability of the random forest model to accurately predict network traffic with minimal error and significantly reduce false positives serves the objectives of this study. Given the limited attention previously garnered in the use of machine learning techniques for predicting threats in the context of the Nigerian oil and gas industry, this research lays a strong foundation for further studies in this critical area.

5.0 CONCLUSION AND FUTURE STUDIES

This study addresses the critical need for enhanced cybersecurity in the Nigerian oil and gas sector by leveraging machine learning techniques to detect and mitigate information security breaches. The research identifies a significant gap in the application of advanced machine learning models within the sector and demonstrates the limitations of traditional information security measures in keeping pace with evolving threats. Through the development and evaluation of the random forest model, the study shows that while initial models were biased due to imbalanced data, applying data balancing techniques improved the model's accuracy, recall, precision, and F1 score. This indicates that with proper data handling, machine learning models can significantly enhance the sector's cybersecurity posture by enabling more accurate threat detection and prevention. These findings underscore the importance of adopting machine learning as part of a comprehensive cybersecurity strategy in the Nigerian oil and gas industry to protect critical infrastructure and ensure the sector's continued contribution to the national economy.

Future research could explore the integration of other ensemble learning approaches, such as boosting algorithms like XGBoost, to enhance model robustness and scalability. While this study focused

https://doi.org/10.4314/njt.v44i3.10

on improving detection accuracy, the integration of blockchain technology could further enhance data integrity and traceability in cybersecurity frameworks. The potential of blockchain in preventing data tampering, as discussed by recent studies, highlights a promising avenue for securing critical infrastructure.

REFERENCES

[1] Ngobe, E. K. (2020). Information technology: "A sustainable competitive advantage trend in Nigerian oil and gas industry". *International Journal of Business & Law Research*, 8(3), 100-108.

[2] Pashchenko, O., Khomenko, V., Ratov, B., Koroviaka, Y., & Rastsvietaiev, V. (2024). "Main Security Threats to Oil and Gas Infrastructure". In Critical Infrastructure Protection in Response to Terrorist Attacks (pp. 109-116). *IOS Press*.

[3] Idowu, O. A., & Madaki, M. (2021). "Cybercrimes and Challenges of Cyber-Security in Nigeria". *International Journal of Sociology and Development*, 3(1).

[4] Stergiopoulos, G., Gritzalis, D. A., & Limnaios, E. "Counterattacks on the oil and gas sector: A survey on incident assessment and attack patterns". *IEEE Access*, 8, 2020, pp. 128440-128475. https://doi.org/10.1109/ACCESS.2020.3007960

[5] Onuntuei, E. (2018). "Safety, Risk, and Reliability of Cyber Network in Oil and Gas Industry". *International Journal of Teaching*, Education, and Learning, 2(2), 81–97. https://dx.doi.org/10.20319/pijtel.2018.22.8197.

[6] Microsoft InfoSec (2023). What is information security (InfoSec)? Microsoft. Retrieved December 13, 2022, from https://www.microsoft.com/enww/security/business/security-101/what-is-information-security-infosec.

[7] Ugwu, C. (2024, November 11). "Nigeria's oil and gas firms grow revenue to N1.031 trillion in 9 months". *Nairametrics* .https://nairametrics.com/2022/11/11/nigerias-leading-oil-and-gas-firms-grow-revenue-to-n1-031-trillion-in-9-months/

[8] Akpan, U. E. (2024). Cybersecurity in the Petroleum Industry: "A Comprehensive Review of Threats and Mitigation Strategies". *Journal of Engineering Research and Reports*, 26(12), 261-270.

[9] Mallick, M. A. I., & Nath, R. (2024). Navigating the cybersecurity landscape: "A comprehensive review of cyber-attacks, emerging trends, and recent developments". *World Scientific News*, 190(1), 1-69.

[10] Tariq, Z., Aljawad, M. S., Hasan, A., Murtaza, M., Mohammed, E., El-Husseiny, A., ... & Abdulraheem, A. (2021). "A systematic review of data science and machine learning applications to the oil and gas industry". *Journal of Petroleum Exploration and Production Technology*, 1-36.

[11] Nwokonkwo, O. C., Samuel, N. U., Eze, U. F., John-Otumu, A. M. (2024). "Machine Learning Framework for Real-Time Pipeline Anomaly Detection and Maintenance Needs Forecast Using Random Forest and Prophet Model". *Automation, Control and Intelligent Systems*, 12(2), 22-34.https://doi.org/10.11648/j.acis.20241202.11

[12] Jambol, D. D., Ukato, A., Ozowe, C., & Babayeju, O. A. (2024). "Leveraging machine learning to enhance instrumentation accuracy in oil and gas extraction". *Computer Science & IT Research Journal*, 5(6), 1335-1357. https://doi.org/10.51594/csitrj.v5i6.1200

[13] Oyewole, O., Adeleye, O., & Akintoye, S. (2021). "Evaluating the cybersecurity risks and challenges in the Nigerian oil and gas sector. *Journal of Information Security and Applications*, p. 60, 102697.

[14] Adeyeye, O. A., & Omogbadegun, Z. O. (2019). Cybercrime and the Nigerian Oil and Gas Industry: "Threats, Impacts, and Mitigation Strategies". *Journal of Information Security*, 10(4), 312-325.

[15] Hanga, K. M., & Kovalchuk, Y. (2019). "Machine learning and multi-agent systems in oil and gas industry applications": *A survey. Computer Science Review*, 34, 100191. doi: 10.1016/j.cosrev.2019.100191.

[16] Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., & Oza, H. (2021). "Application of machine learning and artificial intelligence in oil and gas industry". *Petroleum Research*, 6(4), 379-391.

[17] Obite, C. P., Chukwu, A., Bartholomew, D. C., Nwosu, U. I., & Esiaba, G. E. (2021). "Classical and machine learning modeling of crude oil production in Nigeria": *Identification of an eminent model for application. Energy Reports*, 7, 3497-3505.

[18] Gupta, A., & Shah, M. (2022). "A comprehensive study on artificial intelligence in the oil and gas sector". *Environmental Science and Pollution Research*, 29(50), 50984–50997. https://doi.org/10.1007/s11356-021-15379-z

[19] Adebayo, O. S., & Adetunmbi, A. O. (2021). "Cybersecurity Risk Management Framework for the Nigerian oil and gas sector". *International Journal of Scientific and Technology Research*, 10(7), 348-354.

[20] Umaru, S. B., Inuwa-Dutse, I., & Abdullahi, I. (2021). Artificial Intelligence: "A Paradigm Shift for Enhancing Cybersecurity in the Nigerian Oil and Gas Sector". *International Journal of Advanced Computer Science and Applications*, 12(1), 324-333.

[21] Elijah, O., Ling, P. A., Rahim, S. K. A., Geok, T. K., Arsad, A., Kadir, E. A., & Abdulfatah, M. Y. (2021). "A survey on Industry 4.0 for the oil and gas industry: upstream sector". *IEEE Access*, 9, 144438-144468.

[22] Elijah, O., Ling, P. A., Rahim, S. K. A., Geok, T. K., Arsad, A., Kadir, E. A., & Abdulfatah, M. Y. (2021). "A survey on Industry 4.0 for the oil and gas industry: upstream sector". *IEEE Access*, 9, 144438-144468.

[23] Gupta, P. (2021). Practical Data Science with Jupyter: "Explore Data Cleaning, Pre-processing, Data Wrangling, Feature Engineering, and Machine Learning using Python and Jupyter (English Edition)". *BPB Publications*, *India*.

[24] Gupta, P., Sehgal, N. K., & Acken, J. M. (2024). Machine learning algorithms. In Introduction to Machine Learning with Security: "Theory and Practice Using Python in the Cloud". *Springer International Publishing*. pp. 45-176.

[25] Chan, J. Y. L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). "Mitigating the multicollinearity problem and its machine learning approach": *A Review: Mathematics*, 10(8), 1283.

[26] Alduailij, M., Khan, Q. W., Tahir, M., Sardaraz, M., Alduailij, M., & Malik, F. (2022). "Machine-learning-based DDoS attack detection using mutual information and random forest feature importance method". *Symmetry*, 14(6), 1095.

[27] Sayed, M. A., Sarker, M. S. U., Al Mamun, A., Nabi, N., Mahmud, F., Alam, M. K., ... & Choudhury, M. Z. M. E. (2024). "Comparative Analysis of Machine Learning Algorithms for Predicting Cybersecurity Attack Success: A Performance Evaluation". *The American Journal of Engineering and Technology*, 6(09), 81-91.

[28] Omonayin, E., Akande, O. N., Muhammad, A., & Enemuo, S. (2025). "Evaluating Deep Learning Models for Real-Time Waste Classification in Smart IOT Environment". *Nigerian Journal of Technology*, 44(2), 357-366.

[29] Bayode, O., Aiyelokun, O., Osanyinlokun, O., & Adanikin, A. (2025). "Enhancing Road Crash Prediction: A Comparative Study of Machine Learning Algorithms and Safety Performance Functions on The Lagos-Ibadan Expressway". *Nigerian Journal of Technology*, 44(2), 215-221