

Random Forest Algorithm for Email Threat Detection

Oluwasogo Adekunle Okunade

Department of Computer Science, National Open University of Nigeria, Abuja, Nigeria

Corresponding Author's Email: aokunade@noun.edu.ng

ABSTRACT

Cybercrime manipulation is on the rise, especially email-based threats; traditional security systems are struggling to keep up with the challenges posed by these activities and secure the cyber world. The paper presents a Random Forest algorithm for detecting and preventing malicious email activity. To bridge this gap, the study used a combination of rule-based filtering and machine-learning classification to analyse a dataset of over 10,000 emails, using Python for system deployment and Random Forest for classification, with some flagged as malicious. The Random Forest model was trained and tested using a 60/40 data split and showed impressive results, with an accuracy of 88.04%, an AUC of 0.973, and an F1-score of 0.879, outperforming other model classifiers such as decision trees and confusion matrices. The study suggests that machine learning, particularly Random Forest, can significantly improve email threat detection and help protect cyberspace. These affirm the potential of machine learning, particularly Random Forest, in enhancing email threat detection and forecasting potential victims.

Keywords: Cybersecurity, Threats, Machine Learning, Real-Time Detection, Algorithm,

1.0 INTRODUCTION

The rapid advancement and widespread accessibility of digital technologies have significantly transformed global communication and economic systems, while simultaneously enabling new forms of cybercriminal activity. The internet has become critical infrastructure for both legitimate and malicious operations, facilitating data theft, financial fraud, and system disruption. As reliance on digital communication, particularly email, continues to increase, so too does exposure to cyber threats, making email security an increasingly urgent concern (Naidu & Gowri, 2023). Cybercrime, defined as malicious activities conducted in cyberspace with the intent to cause harm or exploit vulnerabilities, has therefore emerged as a major global challenge affecting both economic development and social stability (Obia, 2021).

The scale and impact of cybercrime are profound, with individuals, organisations, and governments experiencing significant financial and operational losses. Activities perpetrated by cybercriminals, including hacking, phishing, and online scams, have resulted in widespread damage and disruption. Projections suggest that cybercrime could generate global losses of up to \$10 trillion, positioning it among the most economically consequential forms of crime in modern history (Pascal & Nwankwo, 2021). Furthermore, the transnational nature of cyber threats complicates enforcement and mitigation efforts, as attacks often span multiple jurisdictions, thereby necessitating coordinated and adaptive global responses (Safat, Asghar & Gillani, 2021).

Despite ongoing advancements in cybersecurity technologies, persistent vulnerabilities within digital systems continue to be exploited by increasingly sophisticated attackers. Cybercriminals continuously evolve their techniques to circumvent traditional security measures, rendering many conventional defences insufficient. Existing approaches, ranging from legislative frameworks to traditional and hybrid technological solutions, often struggle to keep pace with the dynamic threat landscape (Jabez, 2021). As a result, there is a

growing need for intelligent, automated systems capable of proactively identifying and mitigating cyber threats in real time (Okunade, 2017; Brimah & Singh, 2023; Jasmeen, 2022).

In response to these challenges, the integration of Artificial Intelligence (AI) and Machine Learning (ML) has emerged as a promising approach to enhancing cybersecurity capabilities. Machine learning techniques enable systems to learn from historical data, detect patterns, and adapt to new threats with minimal human intervention (Al-majed & Ibrahim, 2022; Bilen, 2021). Algorithms such as Random Forest have demonstrated strong performance in identifying spam, phishing, and malware, offering improved accuracy and responsiveness compared to traditional methods (Mohler, 2022; Zhang & Farha, 2022). Consequently, AI-driven approaches provide a scalable and proactive framework for detecting and mitigating evolving email-based threats, thereby enhancing the protection of digital communication systems.

2.0 LITERATURE REVIEW

2.1 Concept of Cyber-Crime

Cybercrime encompasses a broad spectrum of illegal activities conducted *באמצעות* digital devices and networked systems, including offences such as fraud, identity theft, data breaches, malware dissemination, and email-based attacks (Al-majed & Ibrahim, 2022). It is generally defined as any criminal activity involving a computer, network, or connected device, whether as the target or as the tool for execution. Cybercriminals may pursue financial gain through activities such as ransomware, online fraud, and payment data theft, or they may aim to disrupt and damage systems directly by deploying malicious software or disabling infrastructure (Mohler, 2022; Cohen & Felson, 2020; Ford, 2020). In some instances, cybercrime operates through hybrid mechanisms, in which systems are initially compromised and subsequently used to propagate malware across networks, thereby amplifying the attack's scale and impact.

From regulatory and conceptual perspectives, cybercrime encompasses a wide array of malicious actions, including

unauthorised data interception, system interference, and violations of intellectual property rights (Hunt, 2022). It is commonly categorised into three dimensions: crimes in which computing systems are the primary target; crimes in which computers function as instruments of attack (such as denial-of-service operations); and crimes in which digital systems act as facilitators or repositories of illicit activity (Al-shehri, Khan, & Yahaya, 2023). While cybercrime represents an extension of traditional criminal behaviour into digital environments, its distinguishing feature lies in the central role of networked technologies in enabling and scaling such activities (Pascal & Nwankwo, 2021). Moreover, cybercrime increasingly targets the “virtual identities” of individuals and organisations—collections of digital attributes stored across databases—highlighting both the growing dependence on digital infrastructures and their inherent vulnerabilities. Despite attempts at anonymity, cybercriminal activities often leave traceable digital footprints, reinforcing the importance of advanced detection and forensic capabilities in modern cybersecurity efforts (Safat et al., 2021; Al-majed & Ibrahim, 2022).

2.2 Concept of Machine Learning (ML)

Machine learning has become a household term in recent years as the concept has moved from science fiction to a key driver of how businesses and organisations process information. With the pace of data creation continuing to grow exponentially, machine learning tools are pivotal for governments and organisations looking to discover patterns, tease out trends, prevent crime, and chart the most profitable path forward.

According to Bilen (2021) and Al-majed & Ibrahim (2022), Machine Learning (ML) is a subfield of Artificial Intelligence (AI) that uses algorithms trained on data sets to create self-learning models that can predict outcomes and classify information without human intervention. Machine learning is used today for a wide range of commercial purposes, including suggesting products to consumers based on their past purchases, predicting stock market fluctuations, creating self-driving vehicles, and translating text from one language to another or audio to text, etc. (Mohammed & Praniitha, 2023; Owen, 2023).

It is important to note that, in common usage, the terms “Machine Learning” and “Artificial Intelligence” are often used interchangeably due to the prevalence of ML in AI today. But the two terms are meaningfully distinct. While AI refers to the general effort to create machines capable of human-like cognitive abilities, ML specifically refers to the use of algorithms and datasets to achieve this.

In the view of Zhang and Farsha (2022), ML is a technique that discovers previously unknown relationships in data by searching potentially very large data sets to discover patterns and trends that go beyond simple statistical analysis. Machine learning uses sophisticated algorithms that are trained to identify patterns in data, creating models. Those models can be used to make predictions and categorise data. Also, it should be noted that an algorithm isn't the same as a model (Sangani & Harooyzarger, 2021; Vitally & Ford, 2023). An algorithm is a set of rules and procedures used to solve a specific problem or perform a

particular task, while a model is the output or result of applying an algorithm to a data set. Before training, you have an algorithm. After training, you have a model.

2.3 Concept of Machine Learning (ML)

Machine learning (ML) is characterised by several core capabilities, including the automatic discovery of patterns, the prediction of outcomes, the generation of actionable insights, and the efficient processing and analysis of large-scale datasets (Tseloni, 2021). These features enable ML systems to learn from historical data and improve performance over time with minimal human intervention. Broadly, ML can be categorised into four principal paradigms, each suited to different problem domains and data conditions. Supervised Learning relies on labelled datasets to train algorithms to map inputs to known outputs, making it particularly effective for classification tasks such as spam email detection. In contrast, Unsupervised Learning operates without labelled data, instead identifying inherent structures or patterns within datasets, often through clustering or anomaly detection, which is widely applied in recommendation systems and fraud detection (Carplan & Keith, 2021; Ford, 2020; Chuknov & Ivanov, 2021).

Complementing these approaches are Semi-Supervised Learning and Reinforcement Learning, which address limitations associated with purely labelled or unlabelled data. Semi-supervised learning integrates both labelled and unlabelled data to enhance model training efficiency, particularly in scenarios where labelled data is scarce or costly to obtain, using iterative techniques such as pseudo-labelling to refine predictions (Liggett & Lee, 2020; Dubber, 2020; Manisha & Kalpana, 2023). Reinforcement learning, on the other hand, is goal-oriented and employs a trial-and-error mechanism in which an agent learns optimal decision-making strategies through feedback in the form of rewards or penalties. This approach is especially suited to dynamic and complex environments, where decisions must adapt to changing conditions over time (Surendran, 2022; Vitaly & Ford, 2023). Collectively, these paradigms demonstrate the versatility of machine learning in addressing a wide range of analytical and predictive challenges.

2.4 Concept of ML Algorithm

Machine learning algorithms form the computational foundation of intelligent systems, enabling the transformation of training data into predictive models that solve complex problems (Brimah & Singh, 2023; Ford, 2020). These algorithms are designed to identify patterns, detect anomalies, and generate predictions across a wide range of applications, including image recognition, fraud detection, and recommendation systems. Among the most widely used techniques are Neural Networks, which simulate interconnected processing nodes inspired by the human brain to extract features from data, and Linear Regression, which models relationships between variables to predict continuous outcomes. Logistic Regression extends this approach to classification tasks involving binary outcomes, while Clustering enables the grouping of similar data points without prior labels. Additionally, Decision Tree models provide interpretable, rule-based

techniques will enable the researcher to assess model performance versus the intended goals. Use of key performance indicators (KPIs) such as accuracy and recall is combined with feature significance evaluation in Random Forest models. To improve the accuracy in crime prediction and local risk assessment, the work combines Machine Learning methods, geographical analysis, and email attack trend monitoring. Following privacy laws and government rules directs the use and access of data. Using best-performing models for predicting criminal acts, such as Random Forest, SVM, and Logistic Regression, the project seeks to create efficient safety promotion programs for the Nigeria Police Force departments.

3.2 Objective of the Design

The system is designed to detect and mitigate email-based cyber threats using a hybrid approach that combines:

Rule-Based Filtering: Quickly screens emails using predefined heuristics, such as suspicious domains, flagged IPs, and risky attachments.

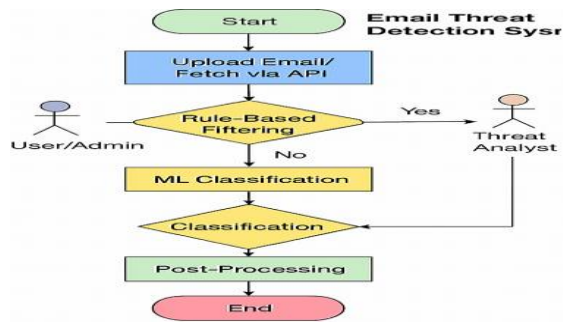
Machine Learning Classification: Uses the Random Forest algorithm to intelligently classify emails as malicious or safe based on extracted features

Main Menu



```
graph TD; User/Admin --> Upload[Upload Email/ Fetch via API]; Upload --> Preprocessing; Preprocessing --> RuleBased[Rule-Based Filtering]; RuleBased --> ML[ML Classification]; ML --> Review[Review & Manage Alerts]; ThreatAnalyst[Threat Analyst] --> ML; Review --> Threat[Threat];
```

134

**Figure 3: Structure of Email Threat Detection System**

A random forest classifier (Manisha & Kalpana, 2023) is well known as an ensemble classification technique used in Machine Learning and data science across various application areas. This method will analyse structured features extracted from email headers, content, and metadata, classify the emails as malicious or benign with high precision, and contribute to a hybrid detection pipeline alongside rule-based filtering. The Random Forest algorithm is trained on a labelled data set, with patterns marked as legitimate or phishing (Reddy K.T, 2023). By leveraging the strengths of Random Forest, the system will achieve greater

accuracy, resilience to noise, and faster response times, making it a reliable tool for proactive cybersecurity defence.

3.3 Data Analysis and Modeling

3.3.1 Data Description

The dataset is part of a malicious email threat intelligence dashboard that tracks and analyses suspicious emails received over 30 days. It contains 100 email entries, of which 70 are flagged as malicious. Each entry includes details such as sender, subject line, timestamp, IP address, threat level, threat score, domain, country of origin, malware type, and any attachments.

Table 1: Data set for email threat analysis

Sender	Subject	Timestamp	Ip_address	Domain	Country	Attachment
donald@account-alert.net	Important Update: New security features	29/07/2025 15:59	147.206.65.150	account-alert.net	Romania	invoice.zip
tonya@notifications.net	Password reset confirmation	10/07/2025 02:51	239.113.103.91	notifications.net	Russia	data.xlsx
jacqueline@account-alert.net	Your package is ready for delivery	09/07/2025 02:22	97.1.146.97	account-alert.net	China	document.pdf
amanda@account-alert.net	Account locked - Immediate action required	01/08/2025 14:06	131.239.22.126	account-alert.net	Germany	
gregory@verify-account.org	Warning: Your computer is infected	21/07/2025 04:15	101.49.152.149	verify-account.org	Romania	document.pdf
kayla@business.net	Shipping confirmation	22/07/2025 20:35	149.217.168.166	business.net	Nigeria	
jamie@account-verification.net	URGENT: Your account will be suspended!	14/07/2025 04:25	234.53.69.46	account-verification.net	Germany	update.exe
christopher@notifications.net	Thank you for your order	28/07/2025 22:52	126.128.224.86	notifications.net	Russia	
willie@security-alert.info	Re: Your recent transaction	19/07/2025 04:16	181.0.225.235	security-alert.info	Romania	invoice.zip
chad@account-alert.net	Urgent: Tax refund pending	14/07/2025 03:39	21.47.140.77	account-alert.net	United Kingdom	photo.jpg.exe

brenda@work.co	Invoice for services	28/07/20 25 02:17	238.88.147.80	work.co	United States	data.xlsx
krista@secure-bank-update.com	Congratulations! You've won \$1,000,000!	22/07/20 25 08:01	197.95.198.219	secure-bank-update.com	Nigeria	

[...10000 more rows...]

The above table shows the dataset presents a detailed snapshot of cyber threat activity through 100 flagged emails collected within 30 days. Each entry includes sender identity, subject line, timestamp, IP address, domain, country of origin, and any attached files. The data reveals a widespread and coordinated use of deceptive email tactics across multiple regions. Domains like account-alert.net, security-check.info, and urgent-update.net appear frequently, suggesting

they are part of sustained phishing or malware campaigns. These domains are often paired with emotionally charged or urgent subject lines- “- “Verify your identity to continue,” “Claim your prize now,” and “Your account will be suspended” - designed to manipulate recipients into opening malicious attachments.

Table 2. Threat Level Classification

Score Range	Threat Level	Description	Action Required
8.0 – 10.0	Critical	Immediate security threat	Block immediately
6.0 – 7.9	High	High probability of malicious	Quarantine & investigate
4.0 – 5.9	Medium	Potentially suspicious	Enhanced monitoring
2.0 – 3.9	Low	Minimal risk	Standard filtering
0.0 – 1.9	Safe	Appears legitimate	Allow with logging

The Email Threat Project uses a Threat Level Classification Table to assign each incoming email a numerical risk score between 0.0 and 10.0. This score is then mapped into one of five categories, Safe, Low, Medium, High, or Critical, which determine how the system responds, such as allowing, flagging, quarantining, or blocking the message. Emails classified as Safe (0.0–1.9) show no signs of malicious

intent, while Low (2.0–3.9) and Medium (4.0–5.9) levels indicate increasing suspicion, including minor anomalies, questionable links, or unusual attachments. High (6.0–7.9) and Critical (8.0–10.0) categories represent strong to severe threats, often involving phishing attempts, spoofed identities, malware, or advanced cyberattacks.

Table 3: Summary of Data Cleaned

Malware Type	Count	Common Attachments	Common Countries	Common Domains	Threat Behaviour
Ransomware	12+	invoice.zip, update.exe	Romania, Nigeria, Russia	account-alert.net, bank-security.co	Encrypts files, demands payment
Botnet	10+	photo.jpg.exe, update.exe	Iran, Ukraine, Brazil	claim-prize.com, secure-bankupdate.com	Controls infected devices remotely
Phishing	9+	update.exe, document.pdf	India, Canada, China	prize-claim.org, verifyaccount.org	Tricks users into revealing credentials
Adware	8+	document.pdf, update.exe	Germany, France, Iran	securitycheck.info, banksecurity.co	Displays ads, tracks user behaviour
Trojan	7+	invoice.zip, update.exe	Romania, Australia	verifyaccount.org, security-alert.info	Disguises as legitimate software

Keylogger	6+	photo.jpg.exe, update.exe	Ukraine, Canada	urgent-update.net, account-alert.net	Records keystrokes to steal data
Spyware	6+	document.pdf, report.pdf	Russia, Germany	secure- bankupdate.com, bank-security.co	Monitors user activity secretly
Rootkit	5+	document.pdf, invoice.zip	China, Pakistan	securitycheck.info, account-alert.net	Hides malware from detection
Virus	5+	update.exe, photo.jpg.exe	Germany, UK	security-alert.info, urgent-update.net	Replicates and damages systems
Worm	5+	photo.jpg.exe, invoice.zip	China, Iran	claim-prize.com, accountverification.net	Spreads across networks automatically

Analysis of the cleaned dataset from the Email Threat Project reveals a broad and complex distribution of email-borne malware, categorized into ten principal types: Ransomware, Botnet, Phishing, Adware, Trojan, Keylogger, Spyware, Rootkit, Computer Virus, and Computer Worm. Each category is examined in terms of frequency, delivery mechanisms, geographic targeting, associated domains, and behavioural signatures. Among these, ransomware emerges as the most dominant threat, frequently delivered through attachments such as compressed archives and executable files, and associated with regions including Romania, Nigeria, and Russia. Its operational pattern, encrypting user data and demanding payment, underscores its high-impact nature. Similarly, botnets and phishing campaigns are widespread, leveraging deceptive file formats and domain impersonation to compromise systems and extract sensitive information, with notable activity in Iran, Ukraine, Brazil, India, Canada, and China.

Other malware types, including adware, trojans, keyloggers, spyware, rootkits, viruses, and worms, demonstrate varying degrees of prevalence and functional specialisation. Adware primarily introduces intrusive advertising and tracking mechanisms, whereas trojans act as covert entry points for broader system compromise. Keyloggers and spyware operate stealthily to capture user activity, while rootkits conceal malicious processes to evade detection. Viruses and worms, by contrast, emphasise replication and propagation across systems and networks. Across all categories, executable files, particularly those masquerading as legitimate updates, constitute the most common delivery vector, highlighting a consistent exploitation of user trust. Additionally, recurring domain patterns and the frequent association of threat activity with countries such as Germany reinforce the transnational nature of email-based cyber threats. Collectively, these findings emphasise the need for comprehensive, data-driven threat detection frameworks that can identify diverse malware

behaviours while accounting for evolving global distribution patterns.

3.6 Random Forest and Feature Selection

This section illustrates the automatic selection process, emphasising the importance of advanced techniques such as random forest and feature selection. Through this approach, crucial input features for understanding any email attack dynamics and output variables for predictive modelling were determined. This integrated approach enhances researchers' ability to derive actionable insights from complex datasets, aiding informed decision-making in analysing and preventing the attack.

3.7 Validation and Testing Procedures

3.7.1 Data Partitioning

First, the dataset was split into two parts: training and testing sets. A common technique called a train-test split was used, where 80% of the data was used to train the models, and the remaining 20% was set aside for testing. This ensures that the models learn from most of the data while still leaving some data unseen to test their performance. Additionally, cross-validation, which divides the data into multiple subsets (folds) and trains the model iteratively on different combinations of these subsets, was chosen to ensure robustness in our evaluation process.

The above figure illustrates the partition sizes for the training and testing sets. In the "Train and test" row, the training partition size is 80%, and the testing partition size is 40%. The labels for these partitions are "Training" and "Testing," respectively. The corresponding values are "1_Training" and "2_Testing." The second row, "Train i and Test," has both partition sizes set to 0, indicating that this partitioning scheme was not used in the current study. By employing these validation and testing techniques, the researcher aims to ensure that the machine learning models are properly evaluated, and their performance is assessed on unseen data, providing a more reliable indication of their generalisation capabilities.

4.0 Results

The Email Threat Tracker dashboard represents a critical analytical and visualisation component within the broader email threat detection framework, enabling real-time monitoring, classification, and response to malicious email activity. By aggregating and processing data across organisational communication channels, the system transforms raw email logs into actionable intelligence for cybersecurity teams. Key performance indicators displayed at the interface, such as 100 total emails scanned, 20 identified as malicious, and 6 classified as high-risk, provide an immediate overview of the threat landscape, illustrating that a significant proportion of communications may contain harmful content. This structured presentation supports rapid situational awareness and prioritisation of defensive actions.

The dashboard further enhances analytical depth through visual and temporal representations of threat data. A categorical distribution of threats shows that the majority fall within the low-risk category, while medium, high, and critical threats comprise smaller but operationally significant proportions. Temporal analysis via a 30-day activity graph reveals distinct spikes in malicious activity, suggesting coordinated campaigns or exploitation of vulnerabilities at specific intervals. Additionally, detailed tabular records of

flagged emails, including sender identity, domain origin, subject line, threat score, and timestamp, facilitate granular investigation of attack patterns such as phishing and spoofing attempts. The inclusion of domain intelligence, particularly through a blacklist of known malicious domains, enables proactive threat mitigation by preventing repeated exposure to identified sources of compromise.

Beyond visualisation, the system integrates machine learning, specifically Random Forest, to automate threat scoring and classification based on email metadata and behavioural indicators. Functional features such as bulk data upload, dataset preview, and filtering tools support scalable analysis and seamless integration with external data sources. Importantly, the system also tracks classification accuracy by identifying false positives, with only 1 misclassified email out of 100, demonstrating high precision. This continuous feedback mechanism enhances model refinement and ensures that legitimate communications are not unnecessarily disrupted. Collectively, the dashboard exemplifies a data-driven, adaptive approach to email security, strengthening organisational resilience against evolving cyber threats.

4.1 Performance Metrics

Table 4: Random Forest model achieved the following:

Metric	Score	What It Means for Email Threats
Accuracy	88.04%	Out of all emails, 88% were correctly classified as either malicious or safe. This shows strong overall reliability.
Precision	0.871	When the model flagged an email as malicious, it was correct 87.1% of the time. This reduces false alarms and helps avoid blocking legitimate emails.

Recall	0.871	The model successfully identified 87.1% of all actual malicious emails. This is crucial for catching threats before they reach users.
F1-Score	0.879	Balances precision and recall. A high F1-score means your model is both accurate and consistent.
AUC	0.973	The model is excellent at distinguishing between safe and malicious emails. AUC close to 1.0 means near-perfect classification ability.

ROC stands for **Receiver Operating Characteristic**. It's a graphical plot that helps you evaluate how well your classification model (like Random Forest) distinguishes between two classes- in your case: Malicious emails (positive class) and Safe emails (negative class).

The ROC curve helps to visualise how well the model balances the goals to catch as many threats as possible

4.2 ROC Curve Analysis

The ROC curve shows the trade-off between:

- **True Positive Rate (TPR)** – also called **Recall**: the proportion of actual threats your model correctly identifies
- **False Positive Rate (FPR)**: the proportion of safe emails your model incorrectly flags as threats.

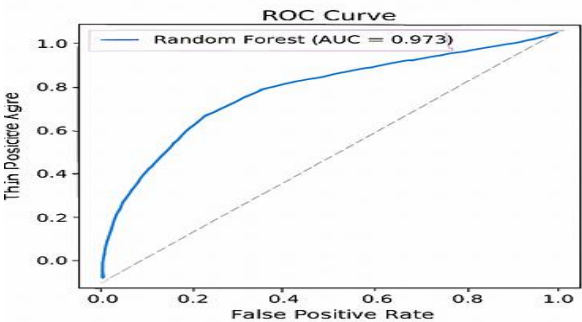


Figure 4: ROC CURVE

The Random Forest model had an **AUC (Area Under the Curve)** of **0.973**, which is excellent. The ROC curve confirms that your Random Forest model is highly effective for email threat detection. It's not just accurate, it's **strategically smart**, catching threats while minimising disruption to safe communications. Ultimately, the Random Forest model isn't just technically sound, it's contextually powerful. It offers a practical, data-driven solution to a real-world problem in Nigeria. With further refinement and larger datasets, it could be deployed in government agencies, banks, or even public awareness platforms to reduce cybercrime.

The Random Forest model outperformed other classifiers. Other models showed lower performance, making Random Forest's superiority (Sharma et al., 2023).

5.0 Discussion

The first research question explored the factors influencing the rise of cybercrime in Nigeria, particularly through email-based threats. The study revealed that the location of the crime, day of the week, gender, and employment status significantly impact cybercrime rates. These variables are now embedded into the Email Threat Project's analytical framework to enhance the profiling and detection of suspicious email activity.

The second and third research questions focused on the role of Machine Learning (ML) in improving threat detection and cybercrime prevention, and on developing a model capable of identifying cyberattacks using diverse factors. This study on the implementation of a Random Forest algorithm yielded impressive results, achieving over 88.041% accuracy, and validated its effectiveness in classifying malicious email patterns and predicting future threats.

These findings align with global crime analytics, which emphasise the importance of demographic, socio-economic, and geographic factors in cybercrime prevalence. The Email Threat detection project reflects this by incorporating contextual metadata, such as sender location, domain origin, and time of email delivery into its threat scoring system.

By applying ML techniques to email metadata, the project supports proactive cybercrime prevention, aligning with emerging research such as Obia (2021). The model's high accuracy suggests strong potential to identify crime trends, enable targeted cybersecurity interventions, and inform data-driven resource allocation.

The project also embraces risk terrain modelling, using email origin and behavioural patterns to identify digital

"hot spots" of cyber activity. This supports community engagement and strategic policing by pinpointing areas and times of heightened threat.

However, the use of ML in crime prediction carries ethical implications. Just as predictive policing can reinforce biases, threat detection models must avoid unfairly flagging emails based on nationality or socio-economic indicators. Ensuring fairness, transparency, and accountability is essential. The Email Threat detector addresses this by: Auditing model outputs for bias, using anonymised and representative datasets, and promoting ethical training for cybersecurity analysts.

Moreover, the quality of email data, like under-reported phishing attempts or incomplete metadata, can affect model reliability. Strengthening public-law-enforcement collaboration and encouraging robust reporting mechanisms are vital for reliable threat intelligence.

Ultimately, the Email Threat Tracker demonstrates how predictive analytics, when combined with community programs and social support, can effectively reduce cybercrime. While data-driven policing is powerful, it must be complemented by social welfare initiatives to address root causes like poverty. Ethical orientation and continuous training in advanced analytics are critical to maintaining public trust and maximising the impact of cybercrime prevention strategies.

6.0 Conclusion

In conclusion, the Email Threat Tracker offers a comprehensive and scalable solution to combat email-based cyber threats, combining advanced technology with ethical oversight to enhance cybersecurity.

The Email Threat Tracker represents a significant advancement in combating cybercrime, particularly email-based threats. By integrating machine learning with contextual crime analytics, the research not only enhances the technical capacity for threat detection but also aligns with global best practices in predictive policing and cybersecurity.

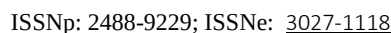
The high accuracy scores achieved by the Random Forest model demonstrate its reliability in identifying and classifying malicious emails. This predictive capability allows for early intervention, reducing the risk of data breaches, financial fraud, and identity theft. Moreover, the inclusion of demographic and socio-economic factors in the model reflects a nuanced understanding of cybercrime dynamics, enabling more targeted and effective prevention strategies.

Future improvements should focus on integrating advanced AI models, such as BERT and Long Short-



Term Memory, to enhance semantic analysis and more accurately detect sophisticated threats. Real-time processing with automated response mechanisms would significantly reduce reaction times, while multilingual support would expand global applicability. Incorporating behavioural analysis and external threat intelligence feeds can improve anomaly

detection and keep the system up to date with emerging threats. Additionally, user-centric filtering, continuous learning through feedback loops, and enhanced dashboard features like predictive analytics will collectively make the system more efficient, adaptive, and precise.



and cyber-crime markets. The Palgrave handbook of

Al-majed, R. & Ibrahim, A. (2022). Using machine learning algorithm for detection of cyber-attacks in cyber-physical systems. *Periodicals of Engineering and Natural Sciences (PEN)*.

Al-shehri, N., Khan, A. and Yahya A., (2023). "Cyber-attack detection framework using machine learning and user behavior analytics," *Computer Systems Science and Engineering*, vol. 44 (2), pp. 1679–1689.

Amjad, R. (2022). Internet-of-Things-Based Suspicious Activity Recognition Using Multimodalities of Computer Vision for Smart City Security", *Security and Communication Networks*, vol. 2. 87-93.

Bilen, A. (2021). "Cyber-attack method and perpetrator prediction using machine learning algorithms," *PeerJ Computer Science*, vol. 7(3).

Brimah, O. and Singh, E. (2023). "Machine learning-based phishing detection from URLs, Expert Systems with Applications, vol. 117, pp. 345-357.

Carplan, J. M., and Keith, J. (2021). Risk terrain modelling: Brokering criminological theory and GIS methods for crime forecasting. *Justice Quarterly*, 28(2), 360-381.

Chuknov, K. and Ivanov, M. (2021). "Machine learning-based mathematical modelling for prediction of social media consumer behaviour using big data analytics," *Journal of Big Data*, vol. 8(13), 123-135.

Cohen, L. E., & Felson, M. (2020). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 588-608.

Dubber, M. (2020). *The Oxford Handbook of Ethics of AI*. Oxford University Press, Incorporated.

Ford, W. F., (2020). A new forecasting approach for crime rates: Adaptive neuro fuzzy inference system. *Journal of statistical and econometric methods*, 7(4), 49-68.

Gunay, D. and Ulfat, S. (2022). National Cyber Security Strategy and the Emergence of Strong Digital Borders. *Connections: The Quarterly Journal*, vol. 19, no. 1, pp. 73–86.

Naidu D. J. and Gowri, M. (2023) Email Phishing Attacks Detection Using Random Algorithm Vol. 4, (8) pp 897-900

Hunt, P., (2022). Evaluation of the Shreveport predictive policing experiment. Santa Monica, CA: Rand Corporation.

Jabez, J. (2021). "Identification of Cyber Threats and Parsing of Data," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 556-564, 2021.

Jasmeen, K. (2022). "Distributed Denial of Service Attacks: A Threat or Challenge," *New Review of Information Networking*, 24 (1) 31-103.

Liggett, R., and Lee, J. R. (2020). The dark web as a platform for crime: An exploration of illicit drug, firearm, CSAM, international cyber-crime and cyber-deviance, pp.91-116.

Manisha, T. and Kalpana, O. (2023). Analysis of Cyber Security Attacks and Its Solutions for the Smart grid Using Machine Learning and Blockchain Methods. *Future Internet* 2023, 15, 83.

Mohler, G. O. (2022). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493), 100-108.

Obia, S.M., (2021). *The Criminal Mind in the Age of Globalisation*. Vij Books India Pvt Ltd.

Okundade, O. A. (2017). Manipulating E-Mail Server Feedback for Spam Prevention. *Arid Zone Journal of Engineering, Technology and Environment*. Vol. 13 (3) Pp. 389–397.

Okunade O.A. (2017). Manipulating E-Mail Server Feedback for Spam Prevention. *Arid Zone Journal of Engineering, Technology and Environment*, Faculty of Engineering, University of Maiduguri, Maiduguri, Nigeria. Print ISSN: 1596-2490, Electronic ISSN: 2545-5818, www.azoje.com.ng Vol. Vol. 13(3). Pp 389-397

Pascal, I. & Nwankwo, O. (2021). Conceptualisation of Cyber-attacks prediction with deep learning. *Cyber-security* 3, 14.

Safat, W., Asghar, S. & Gillani, S. A. (2021). "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques," *IEEE Access*, vol. 9, pp. 70080–70094.

Stalidis, I. H. (2021). Machine Learning for Intelligent Data Analysis and Automation in Cybersecurity: Current and Future Prospects. *Ann. Data. Sci.* (2022).

Surendran, T. (2022). Knowledge in the grey zone: AI and cyber-security. *Digital War*, 1(1-3), 164-170.

Sangani, B. & HarootZarger (2021). Examining deep learning architectures for crime classification and prediction. *Forecasting*, 3(4), pp.741-762.

Tonry, M. (2020). Why crime rates are falling throughout the Western world. *Crime and Justice*, 43(1), 1-63.

Tseloni, A. (2021). Exploring the international decline in crime rates. *European Journal of Criminology*, 7(5), 375-394.

Vitaly, W. and Ford, (2023). Crime rate inference with big data. In *Proceedings of the 22nd ACM International Conference on Knowledge Discovery and data mining* (pp. 635-644).

Wang, D. (2022). Automatic crime prediction using events extracted from Twitter posts. In *International Conference on Social Computing, behavioural-cultural modelling, and prediction* (pp. 231-238).

Zhang, H. & Farha, Y. (2022). Artificial intelligence in cyber security: research advances, challenges, and opportunities. *Artif. Intell. Rev.* 55 (2) 1029–1053.