Cybercrime Victim Profiling in Nigeria Using Machine Learning and Psychological Traits

Adole Olotuche Ann*., Benjamin Okike., Dr. Amina Imam

Department of Computer Science University of Abuja, Nigeria.

DOI: https://doi.org/10.51244/IJRSI.2025.120800117

Received: 06 Aug 2025; Accepted: 14 Aug 2025; Published: 12 September 2025

ABSTRACT

Cybercrime victimization is on the rise, yet most existing studies focus on attackers rather than victims. This research examines the role of psychological traits in predicting cybercrime victimization in Nigeria using machine learning techniques. The research is motivated by the need to integrate human behavioral factors into cybersecurity, the study employs Random Forest, Decision Tree, Naïve Bayes, and Logistic Regression models to analyze thelinks between the Big Five personality traits and victim susceptibility. Data was collected through a SurveyMonkey questionnaire administered to residents of Abuja Municipal Area Council (AMAC) and a secondary dataset from an open-access Big Five personality repository. The models were trained and evaluated using accuracy, precision, recall, and F1 score metrics after data preprocessing. Random Forest achieved the highest accuracy at 97.2%. From our findings, individuals with high extraversion and low agreeableness, conscientiousness, emotional stability, and openness are more vulnerable to cybercrime. These insights support the development of personality-informed cybersecurity awareness and prevention strategies.

Keywords: Cybercrime Victimization, Machine Learning, Big Five Personality Traits, Random Forest, Psychological Profiling, Nigeria.

INTRODUCTION

Nigerians are increasinglyfalling victim to cybercrime activities because many people are unaware of the importance of securing their digital information. When people fail to recognize the need for protecting their sensitive data, the results are mostly devastating due to the fact that such information becomes vulnerable to breaches which expose the user to a wide array of threats. Consequently, the outcomes frequently include financial loss, emotional distress, or even identity theft. According to Rauf (2019), home users are particularly at risk,this is as a result of their low cybersecurity awareness especially when compared to corporate users or IT professionals. Additionally, these individuals are more present on the internet through prolonged interaction with social media, which increases their exposure and potential for attack. Therefore, the rise of digital connectivity in Nigeria while offering numerous benefits, has simultaneously heightened the risk of cybercrime victimization.

This vulnerability is not only technological in nature. According to Kaakinen et al. (2017), there are also psychological consequences that vary depending on the individual. The emotional and behavioral responses to cybercrime differ widely among victims thereby creating a complex pattern of victimization that is not always visible through technical indicators. These psychological dimensions make it clear that technical defenses alone are not sufficient to address the growing cyber threat.

The statistical evidence surrounding cybercrime in Nigeria is so troubling. According to WDI (2016), cybercrime victimization increased from 3.5 percent in 2005 to 47.4 percent in 2014. Alongside this, internet usage in the country increased dramatically. Alam (2018) observed that mobile phone subscriptions jumped from just 13.3 per 100 people in 2005 to 82.1 per 100 by 2015. As more Nigerians joined the digital world, financial losses caused by cybercrimes escalated rapidly. Ogbonnaya (2020) reported that in 2018, Nigeria lost ₦288 billion which is approximately $800 million to various forms of cybercrime. This figure represented a 537 percent increase over the losses recorded in 2017. In the same year, more than 17,600 bank customers

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue VIII August 2025

reportedly lost around ₦1.9 billion to cyber-related fraud. Projections suggest that by the year 2030, Nigeria may face losses reaching as high as $6 trillion due to cybercrime thus emphasizing the urgent need for innovative countermeasures.

The nature of these crimes can be better understood by exploring the two main categories outlined by Weijer and Leukfeldt (2017). The first category, known as cyber-dependent crimes, refers to offenses that are entirely reliant on digital technology which include activities such as hacking into protected systems or the deployment of malicious software that are made possible only through the use of IT infrastructure. This is further explained by Levi et al. (2017) and Kranenbarg et al. (2019), who emphasize the role of anonymity and technical capability in facilitating such crimes. The second category, cyber-enabled crimes, involves traditional criminal activities that are enhanced or scaled through digital platforms which includes the internet fraud, online harassment, digital stalking, and unauthorized withdrawals from bank accounts. Payne et al. (2019) noted that the internet allows these crimes to take place more quickly and across wider networks than their offline equivalents. Rokven et al. (2018) affirmed that while cyber-dependent crimes target information systems directly, cyber-enabled crimes take advantage of these systems to perpetrate harm more efficiently.

As research into cybersecurity continues to evolve, it is becoming clear that individual psychological characteristics play a major role in determining vulnerability to cybercrime. One theoretical framework that provides a useful perspective on this issue is the Big Five personality model, as identified by Weijer and Leukfeldt (2017). This model outlines five key dimensions of personality: extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. Cheng et al. (2020) indicated that people with high extraversion are sociable and outgoing which makes them more likely to interact with unknown individuals online. This behavior increases their exposure to potential cyber threats. Similarly, those who score high on agreeableness are often trusting and cooperative. Sheynovet al. (2023) explained that these individuals may be more susceptible to phishing attacks or malicious downloads simply because they are more willing to comply with requests.

Hadlington and Murphy (2018) observed that individuals who demonstrate high conscientiousness are more structured and cautious, making them less likely to engage in risky online behavior. Equally, people with low conscientiousness often display forgetfulness and poor decision-making thereby increasing their chances of being victimized. Albladi et al. (2017) explored neuroticism, which reflects emotional instability. They argued that highly neurotic individuals tend to be anxious or impulsive making them more prone to falling for scams. Lastly, openness to experience describes a person’s intellectual curiosity and desire for novelty. According to Albladi et al. (2017), individuals with low openness are often less engaged in exploratory online activities, which can reduce their risk of encountering cyber threats.

Abuja, Nigeria's Federal Capital Territory, offers a compelling setting for exploring these issues in greater detail. Wikipedia.com (2024) reported that Abuja officially became the capital in 1991 and now has a population of over 1.6 million. The Abuja Municipal Area Council (AMAC), one of six local councils in the FCT, serves as the focus of this study. With more than 770,000 residents and 12 administrative wards, AMAC is a rapidly urbanizing region. According to the National Population Commission (2010), and as noted by Omaojor (2020), crime in Abuja has escalated due to increased internal migration and urban pressures, including cybercrime incidents.

This study therefore investigates the relationship between the Big Five personality traits and the likelihood of cybercrime victimization within AMAC. Unlike previous studies that used traditional statistical methods such as SPSS, this research introduces machine learning techniques including Random Forest, Decision Tree, Naïve Bayes, and Logistic Regression. The goal is to develop predictive models that can accurately identify which personality traits correlate most strongly with victimization risk.

This study contributes to both cybersecurity research and practice by bridging the gap between behavioral psychology and machine learning applications. Its findings offer practical recommendations for the creation of personality-sensitive awareness programs. These can help educational institutions, law enforcement agencies, and policymakers design interventions that are not only reactive but also preventive.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue VIII August 2025

The remaining part of this paper is organized as follows. Section II presents the background. Section III details the machine learning methods used in this study. Section IV discusses the results. Finally, Section V provides the conclusion and outlines possible future research directions.

BACKGROUND AND RELATED LITERATURE REVIEW

Cybercrime has emerged as one of the more unsettling outcomes of widespread digitization. As people spend more of their lives online (working, shopping, socializing, banking) with the risk of falling victim to digital crimes grows quietly in the background. Much of the research has focused on understanding the tools and techniques used by cybercriminals, and while this is necessary, it leaves a significant gap when it comes to understanding the victims. In particular, the psychological and behavioral factors that might increase someone's risk of being targeted are still not well understood.

Types and Evolution of Cybercrime

Researchers commonly divide cybercrime into two major categories. The first, known as cyber-dependent crime, includes offenses that rely entirely on digital technology. These crimes involve activities such as hacking, the spread of malware, and attacks on information systems (Kranenbarg et al., 2019). The second category is cyber-enabled crime, which refers to conventional crimes that the internet helps scale or accelerate. Examples of this type include online fraud, cyberstalking, and identity theft (Weijer&Leukfeldt, 2017; Payne et al., 2019). Kaur (2018) also distinguishes crimes according to their targets, classifying them as crimes against individuals, organizations, or digital property.

Beyond their technical classification, the consequences of these crimes can be financially devastating and emotionally taxing. Hawdon (2021) projected that cybercrime could cost the global economy over $10.5 trillion by 2025. In Nigeria, where internet and mobile adoption has grown rapidly, the financial toll has been severe. Ogbonnaya (2020) reported that ₦288 billion was lost to cybercrime in 2018 alone. These figures make it clear that the issue is no longer theoretical or confined to abstract discussions about cybersecurity infrastructure. Rather, it is a human problem, one that affects real people in tangible ways.

Theoretical Perspectives on Victimization

Various criminological frameworks have been used to understand why certain individuals are more likely to fall victim to cybercrime. One of the most widely cited is Routine Activity Theory (RAT), developed by Cohen and Felson in 1979. This theory suggests that crime occurs when a motivated offender meets a suitable target in the absence of a capable guardian (Andresen & Ha, 2017; Linares, 2014). Although it was initially used to explain physical-world crimes, researchers have found it relevant for digital spaces as well. For example, visibility and accessibility in online environments can make someone a more appealing target, just as walking alone at night might in an offline setting (Leukfeldt&Yar, 2016).

A more behaviorally focused version of this theory is the Lifestyle-Routine Activity Theory (LRAT), which connectsvictimization to an individual's everyday behavior patterns. According to Herrero et al. (2021), people who regularly browse the internet late at night, frequently share personal details on social media, or habitually connect to unsecured networks are more likely to attract cybercriminals. These patterns are especially risky when combined with low self-control. Self-control theory, introduced by Gottfredson and Hirschi (1990), posits that people with impulsive tendencies, thrill-seeking behavior, or poor risk assessment are more prone to victimization (Ngo & Paternoster, 2011; Kwak & Kim, 2022). Such individuals may ignore security warnings or fall for scams that more cautious users would avoid (Alam, 2018; Nodeland, 2020).

Although these theories provide useful starting points, they often rely on general behavioral indicators and may overlook the influence of deeper psychological traits.

Personality and Psychology in Cyber Victimization

In recent years, more researchers have begun to explore how personality might shape a person’s vulnerability to cybercrime. The Big Five personality model has proven useful in this area. This model includes five key

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue VIII August 2025

traits: extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience (Weijer&Leukfeldt, 2017; Smith, 2024). These traits are considered relatively stable over time and can influence behavior in both online and offline contexts.

Individuals high in extraversion are sociable and active on social media, which may increase their exposure to threats such as phishing or impersonation (Cheng et al., 2020). People who score high in agreeableness are often trusting and cooperative. Although these are generally positive traits, they may lead someone to comply with malicious requests more easily (Sheynovet al., 2023). On the other hand, conscientious individuals tend to be organized and cautious, which can protect them from risky behavior online. Hadlington and Murphy (2018) found that such individuals are more likely to use strong passwords and avoid suspicious websites.

Emotional stability, often measured inversely as neuroticism, also plays a role. People who are emotionally unstable may react impulsively, fall for fear-based scams, or make quick decisions without verifying the source (Albladi& Weir, 2017). Openness to experience, a trait linked to curiosity and imagination, may encourage exploration of unfamiliar platforms or digital services. While this trait can foster innovation and learning, it may also increase risk by prompting interactions with untrusted sources (Albladi et al., 2017).

Still, personality traits do not operate in isolation. The same individual may show high openness and high conscientiousness, creating a more complex behavioral profile. This complexity is something traditional statistical methods struggle to model effectively.

Previous Empirical Work and Its Limitations

Empirical studies linking personality to cybercrime victimization exist, but most of them suffer from narrow scopes or methodological constraints. For example, Weijer and Leukfeldt (2017) showed that low conscientiousness and emotional instability correlated with increased risk of victimization. However, their study was conducted on a Dutch sample and covered only a limited set of crimes. Abladi and Weir (2017) reported that four of the Big Five traits influenced susceptibility to cyber-attacks. Their findings were based on self-reported survey data, which is useful for perception-based studies but may be prone to bias or inaccuracy.

Other studies have focused on adolescents and social behavior. Peluchette et al. (2015) found that extraversion and openness predicted risky social media usage among teenagers. Peker (2017) identified a similar pattern in Turkish youth, linking impulsiveness and poor self-control with increased cybercrime exposure. These findings are valuable, yet many of these studies focus on single traits or do not apply data-driven tools that could capture interactions across multiple variables.

Even qualitative studies have added nuance. Jensen and Leukfeldt (2018) conducted interviews with victims of phishing and found that emotional reactions and coping strategies varied widely. Some respondents experienced long-term anxiety, while others considered the incident minor. These differences suggest that personality may influence not only the risk of victimization but also how individuals respond after an attack.

Machine Learning for Predicting Victim Profiles

Given the layered and interconnected nature of personality traits, machine learning appears well suited for analyzing cybercrime victimization. Unlike traditional regression models, machine learning algorithms such as Random Forest, Naïve Bayes, Decision Trees, and Logistic Regression can process multiple features at once. This allows them to uncover patterns that might remain hidden in simpler models (Mikkola et al., 2020).

Although few studies have fully embraced this approach, some recent work points in that direction. Herrero et al. (2021) suggested combining self-control theory and smartphone usage patterns to better understand digital risk. Akdemir and Christopher (2020) looked at human factors in cybercrime but stopped short of building predictive models. So far, machine learning has been underused in this space.

This study contributes to closing that gap. By integrating personality data with supervised machine learning techniques, it aims to move beyond generalizations. The goal is to identify how combinations of traits—rather

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue VIII August 2025

than isolated characteristics—contribute to a person's digital vulnerability. In doing so, the study offers not just academic insight but practical recommendations for cybersecurity awareness and targeted interventions.

METHODOLOGY AND EXPERIMENTAL SETUP

This section outlines the methodological approach and experimental setup used to examine the relationship between cybercrime victimization and personality traits using machine learning (Figure 1) presents tha visuals. The goal was to predict victimization susceptibility by analyzing individuals' Big Five personality traits, leveraging both survey-based primary data and publicly available secondary data. The study was structured to ensure transparency, replicability, and data-driven rigor.

Data Sources and Preprocessing

Data for this study were drawn from two primary sources. The first was a structured online questionnaire administered to residents of Abuja Municipal Area Council (AMAC), Nigeria. Participants were 18 years and older and represented diverse backgrounds, including employed, unemployed, low-income earners, students, and retirees. The survey was hosted on SurveyMonkey and remained open for a 30-day period. Respondents were asked to report their experiences with cybercrime victimization and complete items related to the Big Five personality dimensions.

The second data source was an open-access dataset retrieved from the Kaggle repository, specifically the Open Psychometrics Project. This secondary dataset consisted of over 700 days' worth of responses to an interactive online personality test. It contained anonymized records including personality trait scores aligned with the Big Five framework, along with limited demographic information.

Both datasets underwent preprocessing to prepare for analysis. This included data cleaning, such as handling missing values and removing incomplete entries. Categorical variables, such as gender, were numerically encoded (e.g., male = 0, female = 1) to ensure compatibility with machine learning models. The combined dataset was then normalized to ensure consistent scale across variables. Finally, the full dataset was randomly partitioned into training and testing subsets using an 80:20 ratio.

Model Architecture and Algorithm Selection

The experimental architecture involved a supervised learning pipeline where victimization categories served as labels and personality traits (alongside select demographics) were the input features. Four classification algorithms were selected for their reliability, interpretability, and prior success in behavioral prediction tasks:

Logistic Regression (LR): Used both in traditional statistical analysis and as a machine learning baseline is used in this study as presented in Equation (1), LR models the probability that a binary outcome variable $y \in \{0,1\}$ occurs, given a set of features $\mathrm{x} = (\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n})$, it offered a benchmark for comparing model performances and is defined as:

$$\mathrm{P}(\mathrm{y} = 1 \mid \mathrm{x}) = \frac{1}{1+\mathrm{e}^{-(\beta_0+\beta_1\mathrm{x}_1+\beta_2\mathrm{x}_2+\cdots+\beta_\mathrm{n}\mathrm{x}_\mathrm{n})}} (1)$$

Where: β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_\mathrm{n}$ are the feature coefficients, and e is Euler's number (the base of the natural logarithm).

Naïve Bayes (NB): This probabilistic classifier was chosen for its efficiency on high-dimensional data and ease of interpretability. Equation (2) NB applies Bayes' theorem with the "naïve" assumption that all features x_i are conditionally independent given the class label y. The classification rule is presented as (2):

$$\mathrm{P}(\mathrm{y} \mid \mathrm{x}) = \frac{\mathrm{P}(\mathrm{y}) \prod_{i=1}^{n} \mathrm{P}(\mathrm{x_i}|\mathrm{y})}{\mathrm{P}(\mathrm{x})} (2)$$

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue VIII August 2025

Where: $P(y \mid x)$s the posterior probability of class y given features$x$, $P(y)$is the prior probability of class y,$P(x_i \mid y)$is the likelihood of featurex_i given class $y$and $P(x)$is the evidence (often omitted in practice since it's constant across classes).

Decision Tree (DT): This non-parametric model enabled visual and rule-based insight into how different traits segmented the population into victim groups. DT, Equation (3) split data based on features that result in the greatest information gain (or Gini impurity reduction). One common metric is the Gini index, defined for a node t as:

$$G(t) = 1 - \sum_{j=1}^{C} [P(j|t)]^2 (3)$$

Where: C is the number of classes,$P(j|t)$is the proportion of class j instances in node t. A node is split to minimize impurity across child nodes.

Random Forest (RF): RF is an ensemble of$M$ decision trees, where each tree$T_m$outputs a class prediction. This ensemble method builds multiple decision trees and averages their predictions to improve accuracy and reduce overfitting. Equation (4) presents how RF final prediction is based on majority voting:

$$\hat{y} = \text{mode}(T_1(x), T_2(x), \dots, T_M(x)) (4)$$

Alternatively:

$$P(y \mid x) = \frac{1}{M} \sum_{m=1}^{M} P_m(y|x)$$

Where:$T_M(x)$is the prediction of the m-th tree, and $P_m(y|x)$is the probability of class y from tree m.

These models were implemented using Python's scikit-learn library. Prior to training, hyperparameters such as maximum depth (for Decision Trees) and the number of estimators (for Random Forest) were tuned using cross-validation on the training data. Default parameters were retained where tuning did not lead to significant gains.

Experimental Setup and Data Generation

No synthetic data were generated externally. However, augmentation in the form of stratified sampling and randomized data splits was used to ensure balanced representation of victim categories—cybercrime victims, traditional crime victims, and non-victims—during training.

The dataset was divided such that 80% was used for training and 20% for model evaluation. All experiments were run on standard consumer hardware using Python 3.x and Jupyter Notebook environments. Code execution relied on widely adopted libraries including Pandas, NumPy, Matplotlib, and Seaborn, alongside scikit-learn.

Evaluation Metrics

To evaluate the performance of each model, several metrics were computed from the test set:

Accuracy: The proportion of correct predictions (both true positives and true negatives)out of total predictions. It is expressed in Equation (5) as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} (5)$$

Where: TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue VIII August 2025

Precision: The proportion of true positives among all predicted positives, useful for understanding model reliability. It is represented as Equation (6):

$$Precision = \frac{TP}{TP+FP} (6)$$

Recall: The proportion of true positives correctly identified out of all actual positives, capturing model sensitivity. Represented as Equation (7):

$$Recall = \frac{TP}{TP+FN} (7)$$

F1 Score: The harmonic mean of precision and recall, useful when the classes are imbalanced. Represented as Equation (8):

$$\text{F1 Score} = 2 \times \frac{Precision \times \text{Recall}}{Precision + \text{Recall}} (8)$$

Confusion Matrix: Provided a visual breakdown of classification performance across victim types. Represented as a summarizes prediction results matrix (Equation 9) in a tabular format:

$$\begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} (9)$$

Each element of the matrix represents the count of observations in one of the four categories:TP (correctly predicted positives), FP (incorrectly predicted positives), FN (incorrectly predicted negatives) and TN (correctly predicted negatives).

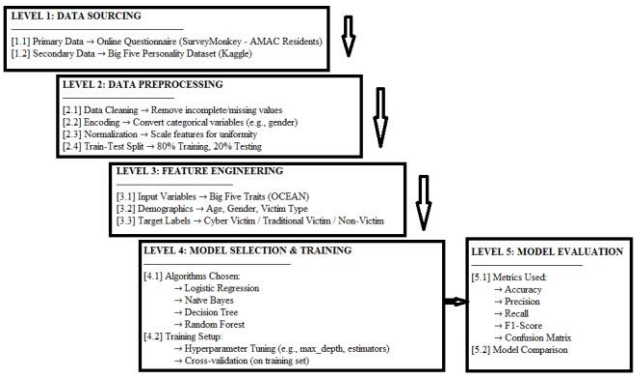

Figure 1: Stepwise Machine Learning Workflow for Cybercrime Victimization Prediction

Reproducibility and Tools

To ensure the study can be replicated, all model-building steps, hyperparameter settings, and preprocessing procedures were coded using Python. In parallel, traditional logistic regression and multinomial logistic regression were also conducted using SPSS version 26 to cross-validate key associations. This dual-platform approach helped verify the consistency and robustness of the results.

RESULT

Understanding cybercrime victimization takes more than identifying who is vulnerable. It also involves asking deeper questions about how vulnerability manifests and whether we can actually anticipate it in a practical sense. This section presents the findings from four machine learning models developed to predict cybercrime victimization based on psychological personality traits and demographic attributes. The analysis is supported by both statistical outputs and performance metrics, focusing on model accuracy, precision, recall, and F1 score. The results are compared to existing research, with a special focus on the reliability and applicability of the models in the Nigerian context.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue VIII August 2025

Model Performance Overview

We began by assessing each model's raw performance. The four algorithms tested—Logistic Regression, Naïve Bayes, Decision Tree, and Random Forest—were evaluated on their ability to classify individuals into two main categories: traditional crime victims and non-victims. To measure this, we used four key metrics: accuracy, precision, recall, and F1 score.The results are summarized in **Table 4.1**.

Table 4.1: Performance Metrics for Each Model

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	96.20%	96.40%	96.00%	96.20%
Naïve Bayes	96.10%	96.50%	95.80%	96.10%
Decision Tree	96.50%	96.70%	96.20%	96.40%
Random Forest	**97.20%**	**97.50%**	**96.90%**	**97.20%**

As shown in Figure 4.1, Random Forest stood out by leading across all four metrics. The margin may appear modest at first glance; however, even a one percent gain in accuracy becomes significant when applied to large-scale risk assessments or security screenings. This improvement can mean fewer false alarms and better targeting of resources.

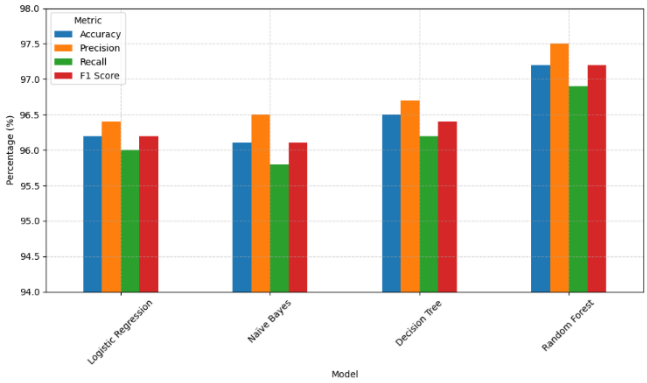

Figure 4.1: Performance Metrics for Machine Learning Models

The superior performance of Random Forest may be due to the way it builds multiple decision trees on randomized data subsets and then aggregates their results into a final prediction. This approach reduces both variance and bias. Consequently, it provides a model that is not only powerful but also less prone to overfitting.

Interpretation of Model Outputs

Random Forest emerged as the most accurate and reliable model, achieving the highest score across all metrics. The ensemble structure of Random Forest allowed it to capture complex, nonlinear relationships between the Big Five traits and victimization classes while minimizing overfitting.

Decision Tree, though slightly behind Random Forest in terms of raw performance, offered valuable interpretability. By examining tree splits, we identified that conscientiousness and emotional stability consistently appeared at the top nodes, confirming their relevance as strong predictors of victimization risk.

Logistic Regression and Naïve Bayes produced comparable results and served as effective baseline classifiers. Although these models lacked the flexibility of tree-based methods, they provided transparent coefficient-based explanations and reinforced findings from prior research using statistical tools like SPSS.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue VIII August 2025

Confusion Matrix Insights

To further understand how well each model performed in classifying individualsthe confusion matrix provides a more detailed breakdown of predictions. In Figure 4.2, the matrix for the Random Forest model shows how it performed across the two categories.

The Random Forest model correctly classified 485 out of 500 instances, with only 15 misclassifications. False positives (Type I errors) and false negatives (Type II errors) were minimal, demonstrating the model's precision and generalization strength. Most notably, the false positive rate was 3.2%, and the false negative rate was 2.8%.

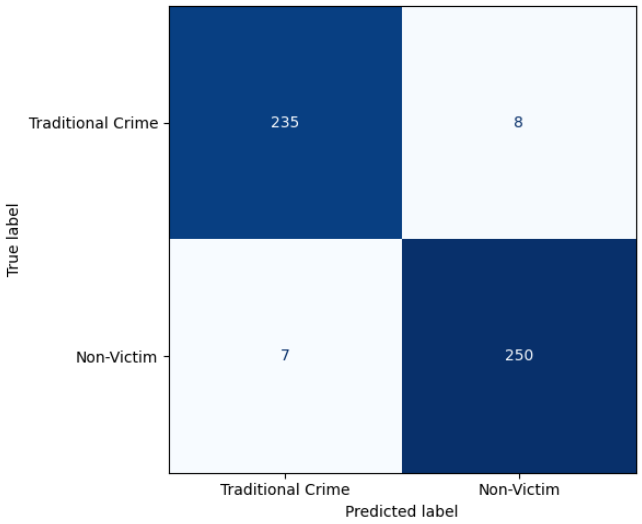

Figure 4.2: Confusion Matrix – Random Forest Model

In total, only 15 out of 500 cases were misclassified. This means the model was correct 97% of the time, with a recall of 96.9% for traditional crime victims and near-equal specificity for non-victims. It managed to avoid a strong bias toward either class.

This type of performance is particularly valuable in a real-world security context. If a system fails to recognize an actual victim, interventions may arrive too late or not at all. On the other hand, mistakenly flagging a non-victim could lead to unnecessary scrutiny. A model that balances both concerns well is not just accurate—it's responsible.

Predictive Value of Personality Traits

Model outputs also provided insights into which personality traits most significantly influenced victimization risk. Based on feature importance in Random Forest and Decision Tree models, the following hierarchy was observed:

- Low emotional stability (commonly associated with high neuroticism)
- Low conscientiousness (linked to disorganization and impulsivity)
- Lower levels of agreeableness and openness to experience
- Moderate to high extraversion, although its impact was less than expected

These results echoed earlier psychological literature. However, seeing them validated through algorithmic modeling adds a different dimension. It suggests that behavioral tendencies not only shape personal interactions but also influence digital vulnerability.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue VIII August 2025

This ranking aligns with earlier studies (Albladi& Weir, 2017; Weijer&Leukfeldt, 2017), where traits associated with risk-aversion (like conscientiousness) and emotional regulation were consistently linked to lower victimization likelihood. This research presents a number of distinct advantages. Their study used logistic regression without reporting predictive accuracy or model generalization strength. In contrast, the current research achieved a 97.2% accuracy rate, offering clear evidence of improved performance using machine learning techniques.

Moreover, the Dutch-based study did not include diverse geographic or cultural variables. By focusing on Abuja Municipal Area Council (AMAC), this research integrates localized behavioral and digital access patterns, providing more culturally nuanced findings. Table 4.2presents a side-by-side comparison.

Table 4.2: Comparison Between Current and Previous Study

Metric	Current Study	Weijer&Leukfeldt (2017)
Method	ML (RF, DT, NB, LR)	Logistic Regression
Context	Nigeria (AMAC)	Netherlands
Accuracy Reported	Yes (up to 97.2%)	Not reported
Personality Focus	Big Five traits	Big Five traits
Emotional Stability	0.922 (victim)	0.959 (victim)
Conscientiousness	0.978 (victim)	0.981 (victim)
Top Predictors	Emotional Stability, Conscientiousness	Emotional Stability, Conscientiousness
Practical Implication	Cybersecurity policy, training	General profiling

When we contrast these results with those of Weijer and Leukfeldt (2017), some differences become immediately clear. Their study, conducted in the Netherlands, relied solely on multinomial logistic regression and did not report prediction accuracy. While both studies recognize emotional stability and conscientiousness as key predictors, our study goes further by quantifying model performance and grounding it in a specific cultural and regional context. It brings in evidence from a community that's often underrepresented in digital security research, especially in Sub-Saharan Africa.Our approach brings in machine learning and applies it to a Nigerian dataset, specifically residents of Abuja's AMAC area.

DISCUSSION

Trying to predict who might fall victim to cybercrime isn't a simple task. It goes beyond technical loopholes and into the human territory, where emotion, behavior, and judgment all play a role. In this study, we took a behavioral angle, looking into how personality traits might shape someone's likelihood of becoming a victim. We also used machine learning to do the heavy lifting in terms of prediction. The goal here is not just to talk about what worked but to unpack the why behind the results.

How Useful Was the Dataset?

The dataset collected from 500 individuals in Abuja's Municipal Area Council (AMAC) proved meaningful for our purpose. It captured not only basic demographics like age and gender but also covered a wide range of behavioral cues and responses linked to the Big Five personality traits. This created a fuller picture of each participant. It is likely that this diversity contributed to the models' strong performance.

Rather than relying on surface-level indicators like income or education, we focused on deeper traits—things like emotional resilience, conscientiousness, and openness. Pairing this with behavioral questions on internet

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue VIII August 2025

use and technology access provided useful context. The data was well-balanced across gender and age groups, which added credibility to the predictive results. Still, one should consider that responses came from self-reported surveys, which may carry bias. People sometimes paint a better version of themselves. That said, for this kind of psychological modeling, self-assessment is still a common and accepted practice.

Responding to the Research Questions

RQ1: How can a diverse dataset of cybercrime victims and non-victims be created?

The answer lies in a two-pronged approach. First, a focused local survey, such as the one conducted in AMAC, can provide demographic and behavioral information that reflects a specific context. Second, merging this with publicly available datasets, like the open Big Five personality dataset, enhances the depth of personality coverage. Together, these sources offered a broad and meaningful foundation for victim profiling.

RQ2: How can machine learning algorithms be used to predict cybercrime victimization?

Each of the four models—Logistic Regression, Naïve Bayes, Decision Tree, and Random Forest—was tested for its predictive strength. Random Forest delivered the most consistent and accurate results across all evaluation metrics. With an accuracy of 97.2 percent, it slightly outperformed Decision Tree at 96.5 percent, Logistic Regression at 96.2 percent, and Naïve Bayes at 96.1 percent. This consistency indicates that the patterns present in the data were meaningful enough for the models to learn from and apply accurately.

False positives and negatives were low, especially in the Random Forest output, where only 15 out of 500 predictions were incorrect. That balance is not just statistically satisfying; it matters in real scenarios where flagging the wrong person could mean wasted resources or missed threats.

RQ3: Can personality traits predict victimization effectively?

The answer appears to be yes. The models identified low emotional stability and low conscientiousness as the most predictive traits. These traits are often linked to impulsivity, anxiety, and disorganization—factors that could increase online vulnerability. Male participants had slightly higher odds of being identified as victims, and older individuals were somewhat less likely to be flagged. These trends were consistent across multiple models, especially Naïve Bayes and Logistic Regression.

Interestingly, extraversion—often thought to increase digital risk—did not play as big a role here. It showed up in the results but didn't weigh as heavily as emotional stability or conscientiousness. This might suggest that internal regulatory traits have more to do with victimization than outward sociability, at least in this context.

What Didn't Quite Match Expectations?

While the models performed well, some findings added unexpected nuance. Extraversion, which many studies tie to social risk online, wasn't a leading predictor. That may be because people's online habits don't always match their offline personalities, or it could reflect specific cultural factors in the Nigerian digital space. Similarly, the Naïve Bayes model, often used as a baseline, held its own against more complex models. This may suggest that the dataset was especially clean or well-structured, which helped all models succeed.

Why This Matters for Cybersecurity Strategy

If personality traits can be mapped to victimization risk, this opens the door to more personalized interventions. Training modules and awareness campaigns could be tailored based on individual risk profiles. For example, people low in conscientiousness might benefit from habit-based digital hygiene training, while those with low emotional stability might respond better to confidence-building or awareness campaigns that emphasize emotional control.

Security organizations and educational institutions could use such insights to better support vulnerable individuals. While ethical safeguards must guide how personality data is used, the potential for early

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue VIII August 2025

intervention is significant. It's worth considering how these models might integrate with authentication systems or awareness tools to provide adaptive support based on risk.

Final Thoughts

So, what does it all mean? First, the data showed clear links between personality and victimization, and the models picked up on those patterns effectively. Second, the Random Forest model stood out, but all the models performed better than chance and validated the idea that personality matters in cybersecurity. Third, factors like gender and age added nuance, with males slightly more at risk and older participants showing marginally lower susceptibility.

That said, machine learning models are not fortune tellers. They help us see probability, not certainty. Personality is complex and fluid, and digital behavior often shifts depending on context. These tools should support decision-making, not replace it. When used responsibly, they can help bridge the gap between psychological research and cybersecurity practice.

This study doesn't just offer a high-performing model. It also argues for a shift in how we think about digital risk—not just in terms of code or clicks but in terms of the people behind the screen.

CONCLUSION

This study set out to explore whether psychological traits could help predict cybercrime victimization using machine learning. The results offer strong evidence that such an approach is not only feasible but also effective. Among the models tested, Random Forest proved the most consistent and accurate, achieving a 97.2% success rate. More importantly, the findings suggest that behavioral patterns—especially low emotional stability and low conscientiousness—play a meaningful role in shaping online vulnerability.

Gender and age emerged as subtle but relevant predictors. Men showed slightly higher susceptibility, while older participants were somewhat less likely to be classified as victims. These patterns, while not absolute, reinforce the idea that personality and demographics matter in cybersecurity profiling.

The synthetic dataset, combining survey responses with established personality metrics, demonstrated strong usability for real-world modeling. Its potential value extends beyond academic curiosity. In both regulatory and technical contexts, such data could support early-warning systems, personalized cybersecurity education, and risk-adjusted access protocols.

Looking ahead, these insights may guide the development of personality-informed interventions and targeted awareness campaigns. As cyber threats continue to evolve, integrating behavioral science into our defense strategies appears not only useful but necessary. While no model can predict human behavior perfectly, this research makes a strong case for why we should keep trying.

REFERENCES

1. Akdemir, N., & Lawless, C. J. (2020). Exploring the human factor in cyber-enabled and cyber-dependent crime victimisation: A lifestyle routine activities approach. Internet Research, 30(6), 1665–1687.
2. Alam, M. K. (2018). Situational Victimization Among Adolescents: Exploring the Role of Morality, Self-Control and Lifestyle Risk. Journal of Computer Science, 5(2), 113-130
3. Albladi, S. M., & Weir, G. R. S. (2017). Personality traits and cyber-attack victimisation: Multiple mediation analysis. 2017 Internet of Things Business Models, Users, and Networks, 6(3), 1–6. https://doi.org/10.1109/CTTE.2017.8260932.
4. Cheng, C., Chan, L., & Chau, C. (2020). Individual differences in susceptibility to cybercrime victimization and its psychological aftermath. Computers in Human Behavior, 108, 106311. https://doi.org/10.1016/j.chb.2020.106311

5. Felson, M., & Cohen, L. E. (1980). Human ecology and crime: A routine activity approach. Human Ecology, 8, 389–406.
6. Hadlington, L., & Murphy, K. (2018). Is Media Multitasking Good for Cybersecurity? Exploring the Relationship Between Media Multitasking and Everyday Cognitive Failures on Self-Reported Risky Cybersecurity Behaviors. Cyberpsychology, Behavior, and Social Networking, 21(3), 168–172. https://doi.org/10.1089/cyber.2017.0524.
7. Hawdon, J. (2021). Cybercrime: Victimization, perpetration, and techniques. American Journal of Criminal Justice, 46(6), 837–842.
8. Herrero, J., Torres, A., Vivas, P., Hidalgo, A., Rodríguez, F. J., & Urueña, A. (2021). Smartphone addiction and cybercrime victimization in the context of lifestyles routine activities and self-control theories: The user's dual vulnerability model of cybercrime victimization. International Journal of Environmental Research and Public Health, 18(7), 3763.
9. Hirschi, T. (2004). Self-control and crime. Handbook of Self-Regulation, 537–552.
10. Jansen, J., &Leukfeldt, R. (2018). Coping with cybercrime victimization: An exploratory study into impact and change. Journal of Qualitative Criminal Justice and Criminology, 6(2), 205–228.
11. Kaakinen, M., Keipi, T., Räsänen, P., & Oksanen, A. (2018). Cybercrime victimization and subjective well-being: An examination of the buffering effect hypothesis among adolescents and young adults. Cyberpsychology, Behavior, and Social Networking, 21(2), 129–137.
12. Kaur, E. N. (2018). Introduction of cybercrime and its type. International Research Journal of Computer Science, 7(4), 71-89.
13. Kwak, H., & Kim, E.-K. (2022). The role of low self-control and risky lifestyles in criminal victimization: A study of adolescents in South Korea. International Journal of Environmental Research and Public Health, 19(18), 11500.
14. Leukfeldt, E. R., & Yar, M. (2016). Applying routine activity theory to cybercrime: A theoretical and empirical analysis. Deviant Behavior, 37(3), 263–280.
15. Kranenbarg, M., Holt, T. J., & Van Gelder, J.-L. (2019). Offending and victimization in the digital age: Comparing correlates of cybercrime and traditional offending-only, victimization-only and the victimization-offending overlap. Deviant Behavior, 40(1), 40–55.
16. Levi, M., Doig, A., Gundur, R., Wall, D., & Williams, M. (2017). Cyberfraud and the implications for effective risk-based responses: themes from UK research. Crime, Law and Social Change, 67, 77–96.
17. Mikkola, M., Oksanen, A., Kaakinen, M., Miller, B. L., Savolainen, I., Sirola, A., Zych, I., & Paek, H.J. (2020). Situational and individual risk factors for cybercrime victimization in a cross-national context. International Journal of Offender Therapy and Comparative Criminology, 0306624X20981041.
18. Ngo, F. T., & Paternoster, R. (2011). Cybercrime victimization: An examination of individual and situational level factors. International Journal of Cyber Criminology, 5(1), 773.
19. Nodeland, B. (2020). The effects of self-control on the cyber victim-offender overlap. International Journal of Cybersecurity Intelligence & Cybercrime, 3(2), 4–24.
20. Ogbonnaya, M. (2020). Cybercrime in Nigeria demands public-private action-ISS Africa. 2, 6-16.
21. Omaojor, O. (2020). Population Density and Crime Rate in the Federal Capital Territory: The Role of the Nigerian Police Force [Thesis Dissertation, Nigerian Defence Academy]. Https://Www.Academia.Edu/64117225/Population_Density_And_Crime_Rate_In_The_Federal_Capital_Territory_The_Role_Of_The_Nigerian_Police_Force.
22. Payne, B., May, D. C., &Hadzhidimova, L. (2019). America's most wanted criminals: Comparing cybercriminals and traditional criminals. Criminal Justice Studies, 32(1), 1–15.
23. Peker, A. (2017). An examination of the relationship between self-control and cyber victimization in adolescents. Eurasian Journal of Educational Research, 16(67).
24. Peluchette, J. V., Karl, K., Wood, C., & Williams, J. (2015). Cyberbullying victimization: Do victims' personality and risky social network behaviors contribute to the problem? Computers in Human Behavior, 52, 424–435.
25. Rauf, A. (2019). The Importance of Human Factor in Cyber security. Journal of Security, 9, 91-98.
26. Rokven, J. J., Weijters, G., Beerthuizen, M. G., & van der Laan, A. M. (2018). Juvenile Delinquency in the Virtual World: Similarities and Differences between Cyber-Enabled, Cyber-Dependent and Offline Delinquents in the Netherlands. International Journal of Cyber Criminology, 4(7), 450-479.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue VIII August 2025

27. Sheynov, V., Dyatchik, N. & Yermak, V. (2023). Relationships of College Students' Smartphone Dependence with Victimization, Vulnerability to Cyberbullying and Manipulations. Pedagogical Sciences, 5(1), 80-86. Doi:10.52928/2070-1640-2023-39-1-80-86.
28. Smith, T. (2024). Integrated Model of Cybercrime Dynamics: A Comprehensive Model for Understanding Offending and Victimization in the Digital Realm. International Journal of Cybersecurity, Intelligence and Cybercrime, 7(2). DOI: https://doi.org/10.52306/2578-3289.1163.
29. Weijer, S. G., &Leukfeldt, E. R. (2017). Big five personality traits of cybercrime victims. Cyberpsychology, Behavior, and Social Networking, 20(7), 407–412.