International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 13– No.4, August 2026 – www.ijais.org

Cybercrime Risks Prediction Model for Secure E-Banking Services

Elizabeth Ufot
Department of Information Technology, National Open University of Nigeria

Saviour Inyang
Department of Cybersecurity, Federal University of Technology, Ikot Abasi, Nigeria

ABSTRACT
The adoption of e-banking in has led to significant growth in digital transactions, but it has also increased the vulnerability of the sector to cyber threats. These threats range from phishing and malware attacks to more sophisticated techniques like Advanced Persistent Threats (APTs) and Distributed Denial of Service (DDoS) attacks, which has caused the financial sector to experience substantial losses. As such, traditional cybersecurity measures are proving insufficient to mitigate these risks. This study proposes a cybercrime risk classification and prediction model based on machine learning algorithms, including Decision Tree and Logistic Regression, to enhance cybersecurity risk assessment in e-banking services, to enhance e-banking security classification. Meanwhile A structured questionnaire was designed and administered for data collection, where data preprocessing techniques such as missing values imputation, data normalization, and feature selection to prepare the data for machine learning models was carried out. The dataset was divided into 80% training and 20% testing, for model training and evaluation. Performance metrics, including accuracy, precision, recall, and F1 score, were calculated. The Decision Tree classifier achieved an accuracy of 66.3%, with precision, recall, and F1 scores of 66.8%, 66.2%, and 66.0%, respectively. In comparison, the Logistic Regression classifier obtained a lower accuracy of 56.7%, with corresponding precision, recall, and F1 scores of 56.4%, 56.2%, and 55.9%. The results indicate that the Decision Tree model provides a more effective classification of cybercrime risk levels, offering, proactive approach to identifying and mitigating emerging threats in digital banking sector.

Keywords
Cybersecurity, E-banking, Cybercrime, Cyber-Attack, Machine Learning, Security

1. INTRODUCTION
In digital economy, the financial services sector has rapidly embraced e-banking, transforming the way financial transactions are conducted worldwide. E-banking includes a variety of online financial services, such as internet banking, mobile banking, and online payment systems, offering customers the convenience of accessing banking services anytime and anywhere[1]. According to a report by the World Bank, over 3.5 billion people globally have adopted some form of e-banking [2], indicating its widespread acceptance and integration into everyday activities. This digital revolution has enabled unprecedented levels of financial inclusion, particularly in regions with previously limited access to traditional banking infrastructure. While e-banking has brought numerous benefits, it has also exposed the financial sector to significant cyber threats[3]. Cybercrime in e-banking encompasses a range of malicious activities aimed at exploiting the vulnerabilities of digital banking systems to gain unauthorized access to sensitive information, steal funds, or disrupt services[4]. The increasing sophistication of cybercriminals poses a substantial challenge to the security of e-banking systems. Global cybercrime damages are expected to reach $10.5 trillion annually by 2025, highlighting the growing scale and impact of digital threats [5]. Cybercriminals employ a variety of techniques, including phishing, malware, ransomware, and social engineering, to target financial institutions and their customers[6]. These attacks can lead to significant financial losses, data breaches, and damage to the reputation of affected institutions. As cybercriminals continue to refine their tactics, it is imperative for financial institutions to adopt more sophisticated and proactive approaches to cybersecurity. Traditional cybersecurity defenses, which predominantly rely on signature-based detection and reactive measures, are increasingly proving inadequate in the face of modern, dynamic cyber threats. These methods often fail to detect new and evolving types of cyber-attacks, allowing cybercriminals to exploit previously unknown vulnerabilities[7], [8]. The integration of artificial intelligence (AI) and machine learning (ML) into cybersecurity offers a promising solution to these challenges. AI and ML technologies can analyze vast amounts of data to identify patterns, detect anomalies, and predict future threats with greater accuracy and speed than traditional methods[9]. Moreover, the volume and complexity of data generated by e-banking transactions make it challenging for conventional systems to identify and respond to threats in real time. The problem is further compounded by the lack of a unified and intelligent system capable of integrating various security measures to provide comprehensive protection against cyber threats. Current approaches are often siloed and reactive, focusing on specific types of threats rather than offering a holistic and proactive defense mechanism. Given the diverse and evolving nature of cyber threats, there is a pressing need for an intelligent cybercrime classification and risk prediction model specifically tailored to e-banking environment. This model would not only enhance the detection and classification of existing and emerging cyber threats but also predict potential risks, enabling financial institutions to take preemptive measures to mitigate these threats.

2. LITERATURE REVIEW
2.1 Cyber Threats Targeting E-Banking Services
Cyber-attacks against electronic banking services include many types of illegal actions performed to gain access to personal information of bank customers, disrupt banking systems, or destroy their reputation[10]. With the use of the internet,

mobile applications, websites, APIs, and other payment systems, there are multiple ways to which cyber-attack can be launched. The most common method of cyber-attack is phishing, where adversaries disguise themselves as banks through emails, SMSs, phishing websites, and mobile phishing attacks to get the victim's login, password, credit card number, one-time password, and others. Often, phishing is combined with more advanced methods of attack like adversary-in-the-middle, which allows attackers to hijack login sessions and bypass some weak MFA mechanisms by stealing session cookies[11]. Malware attacks are also another major cyberattack kind, especially banking Trojans, remote access trojans, overlays, and SMS interceptors[12]. These types of malwares get installed on customer computers through the use of malicious links or fake updates of applications, or malicious files attached to emails. With the emergence of mobile banking, more attacks have appeared due to the fact that mobile phones contain applications used for banking activities and other personal information which makes them vulnerable and suitable for attack[13]. Man-in-the-browser attacks represent a great threat because they work in the browser of the client, which allows attackers to change transaction details once the customer has already logged into his/her account and provided his/her login and password. However, e-banking systems can become victims of API-focused attacks that target backend systems, including mobile apps, websites, payment systems, and other services linked through APIs. Weak API authentication, poor input validation, and overexposure of APIs can be used by attackers to enumerate user accounts, misuse transactional APIs, or obtain large amounts of data. As a result, security of applications has become an important issue for modern banks, as they use cloud platforms and financial ecosystems to provide their services. One more risk is represented by supply chain compromises, as vendors can deliver software, cloud services, identity management, support, and analytical tools. Vulnerability of just one partner can be leveraged to gain access to the bank's systems, as attackers are often looking for the easiest way into the network.

2.2 Cybercrime Risk Prediction and Mitigation for Secure E-Banking Services

Risk prediction and mitigation of cybercrime associated with secure electronic banking services involve approaches and strategies that are focused on identifying the possibility of cyberattacks[14], understanding how such cyberattacks may affect the business and decreasing the possibility of compromising the customer, transaction, or bank. Cybercrime mitigation in banking institutions is very relevant since banking institutions use web portals, apps, application programming interface, clouds, and third-party services[15]. All these elements increase the number of channels that may be used to conduct such cyber activities as phishing, account takeover, malware infection, DDoS attacks, ransomware, and fraud. Today, the machine learning methods of are more frequently used to prevent cybercrimes. In banking industries, models including autoencoders, logistic regression, reinforcement learning, support Vector Machine , Decision Tress, Bayesian and deep sequence methods have been proposed for anomaly detection, fraud classification, and DDoS prevention[16], [17]. The advantage of this approach is that it shifts e-banking protection from reactive incident response to proactive risk anticipation. Mitigation involves turning a prediction of risk into action taken. If threat is detected in a transaction, session, or account, some actions might include the use of step-up authentication, pausing the transaction or session, deferring the transfer, conducting out-of-band verification, or sending the threat to a human analyst for further investigation[18]. In the case of wide infrastructure attack threats, some actions to take might be rate limiting, bot filtering, WAF configuration, DDoS cleaning, network segmentation, and isolation of the endpoint in question[18]. An effective e-banking mitigation plan should have strong authentication and access control measures. All the methods for safeguarding e-banking services, includes phishing resistant multifactor authentication, device binding, behavioral biometrics, certificate-based authentication, and token protection, which will help to lower the risks of account takeovers. Furthermore, model robustness and governance also play a crucial role in mitigation of cyber-attack in e-banking services[19]. The predictive systems can be hacked by data poisoning, adversarial samples or manipulation of the fraud feedback labels especially if organizations rely heavily on automation. Studies on the interaction between artificial intelligence and cybersecurity have shown that the machine learning algorithms used in the prediction processes are vulnerable to evasion and poisoning attacks[20], [21], [22], [23] and thus need to be developed with security, trustworthiness, resilience and robustness in mind. Therefore, it is important for the banks to perform regular validation of their models, detect data drift, conduct human intervention in decision-making and utilize explainable AI in explaining why the particular transaction was considered suspicious.

3. METHODOLOGY

The methodology for this study involves the following steps:

3.1 Data Collection

The primary mode of data collection was used in this research. Data was collected through the use of questionnaire instrument. A sample of the questionnaire is presented in Figure 2. below. The Dataset compiled from the questionnaire consists of 104 observations and 26 predictor variables. A snapshot of the raw data compiled from the questionnaires is presented in figure 2. After the gathering of the information from the questionnaire, the data was uploaded to Kaggle[24] so that it can be publicly accessible not only for this study but for researchers researching in the same field or domain.

3.2 Data Preprocessing

In order to extract information and inferred meanings from the data and gain a better understanding of the events that produced the data, preprocessing of data include transforming the raw data into a format and structure that may be appropriate for these purposes[25], [26]. Data preparation employs well-established statistical tools; some of these methods have been utilize in this study to meet the requirements of the datasets. Additionally, the methods used in this study to preprocess the raw data into a format that can be used for machine learning processes include dataset analysis, data cleaning; there were no missing values in the dataset or entries, variable transformation; the variable transformation technique is used to encode categorical variable into numeric value. In this study, a data transformation method was used. This method simply extracts all the unique categorical values and replaced them with a number. The transformed variable is therefore expressed as

$$x_{transformed} = x_1, x_2, x_3 \dots x_n \quad (1)$$

Where x_i is the numerical value of the categorical label.

Again, data normalization or standardization is a method for transforming data so that its mean and standard deviation are both equal to one. Given that some machine learning methods

are sensitive to the size of the input data, this is frequently done to guarantee that all features in a dataset are on an equivalent scale. Data can be standardized by dividing each centered value by the standard deviation after deducting the mean from each value. This type of standardization is known as the z-score which was utilized in this study and is expressed as

$$z = \frac{x - \bar{x}}{std} \quad 2$$

Furthermore, feature selection is a data preprocessing technique used in determining the most importance predictor variables having the highest influence on the dependent variable. The chi-square technique was used to examine the influence of each independent variable on the dependent variable and the response is ranked based on the feature variables using the $p - values$ of the chi-square statistical test. The corresponding score is expressed as $-log(p)$.

3.3 Model Formulation

Two supervised learning models (Decision tree and Logistic regression) are used in this study. These models are trained using the preprocessed dataset discussed in the previous sections. Details of the model formulation is presented as follows;

3.3.1 Decision Tree

Decision tree uses impunity measures for feature splitting into nodes. The Gini impunity measure was used in this study. For a node with classes $C_1, C_2, \ldots, C_K$, the Gini Impurity G is defined as:

$$G = 1 - \sum_{k=1}^{K} p_k^2 \quad 3$$

Where p_k is the probability of class C_k at the node p_k proportional of samples of the C_k in the node. C_k is the total number of classes in the dataset. To determine the best split, we calculate the Information Gain (IG), which measures the reduction in impurity achieved by a split. For a feature X with a split point s that divides the data into two subsets D_{left} and D_{right}, the information gain IG for the Gini impurity measure is expressed as:

$$IG(X, s) = I(D) - \left(\frac{|D_{\text{left}}|}{|D|} I(D_{\text{left}}) + \frac{|D_{\text{right}}|}{|D|} I(D_{\text{right}}) \right) \quad 4$$

where:

$I(D)$: Impurity of the parent node before the split.

D_{left} and D_{right} : Data subsets after the split.

|D|: Total number of samples in the node.

$|D_{\text{left}}|$ and $|D_{\text{right}}|$: Number of samples in the left and right nodes, respectively. The split point s that maximizes $IG(X, s)$ is chosen for that feature. Considering the binary classification problem addressed in this study with two classes, H_1 and L_1, and evaluating a split of a feature such as age at split point s, the classifier can be model using the Gini for the parent node given that the parent node has p_1 proportion of the high risk H_1 class samples and p_2 proportion of low risk L_1 class samples, where the parent is expressed as follows

$$G_{parent} = 1 - (p_1^2 + p_2^2) \quad 5$$

3.3.2 Logistic Regression

The key element that logistic regression technique employed is the logistic response function and the logit in which we map a probability score (0-1) to a more expansive scale suitable for linear modeling. The first step is to consider the outcome variable not as classes but as a probability p with the label '1'. Naively, it might be tempted to model p as a linear function of the predictor variables given as: $p = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q$, however, fitting this model as such does not ensure that p will always have the values of between 0 and 1 as required from a probability. Instead, we model p by applying a logistic response or inverse logit function to the predictor variables given as:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q)}} \quad 5$$

The equation 5 transformation ensures that the p stays between 0 and 1. To get the exponential expression out of the denominator, we consider odds instead of probabilities. Odds are the ratio of "successes" (1) to "nonsuccesses" (0). In terms of probabilities, odds are the probability of an event divided by the probability that the event will not occur. For example, if the probability that of risk will be high, that is above 0.5, the probability of "low risk" is (1 – 0.5) = 0.5, and the odds are 1.0:

$$Odds(Y = 1) = \frac{p}{1 - p} \quad 6$$

We can the obtain the probability from the odds using the inverse odds function:

$$p = \frac{Odds}{1 + Odds} \quad 7$$

The logistic response function can then be combined with the odds formulation, shown earlier, to get:

$$Odds(Y = 1) = e^{\beta_0} + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q \quad 8$$

Finally, taking the logarithm of both sides, we get an expression that involves a linear function of the predictors:

$$\text{Log (Odds (Y } = 1)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q \quad 9$$

The log-odds function, also known as the logit function, maps the probability p from (0, 1) to any value $-\infty, +\infty$ a linear model is used to predict the probability, which can subsequently be mapped to a class label using a specified cutoff rule, any record with a probability greater than the cutoff is classified as a 1.

3.4 Model Performance Evaluation

Model performance evaluation method aims to determine which model generates the most accurate and practical prediction. Several techniques were employed to assess the effectiveness of the categorization algorithms employed in this study, including the following:

i. Accuracy: This is the most basic metric and represents the number of correct predictions made by the model divided by the total number of predictions. It is expressed mathematically as:

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad 10$$

ii. Precision and recall: These metrics are used when the goal is to minimize false positives or

false negatives. Precision is the ratio of true positives to the total number of predicted positives, while recall is the ratio of true positives to the total number of actual positives. The mathematical representation of the precision and recall are expressed in the following formulas:

$$Precision = \frac{T_P}{T_P + F_p} \quad 11$$

$$Recall = \frac{T_p}{T_n + F_p} \quad 12$$

iii. F1 score: This is a weighted average of precision and recall, and is commonly used when the goal is to balance the trade-off between precision and recall. It is expressed as:

$$F1_{score} = 2 \times \frac{precision \times recall}{precision + recall} \quad 13$$

iv. AUC-ROC: This metric is used to evaluate binary classification models and represents the area under the receiver operating characteristic curve. It measures the model's ability to distinguish between positive and negative classes.

v. Confusion Matrix: The number of precise and unreliable predictions is categorized in a table called the confusion matrix. The effectiveness of a classification model when applied to a set of test data for which the true values are known is usually summarized in a confusion matrix. True positives and negatives are included in the matrix. Additionally, to assess the ratio of true negatives to false negatives and true negatives, or selectivity. The confusion matrix table reveals how many relevant samples the classifier has identified and how many of them have been accurately identified. This recall ratio provides a measurement of inclusiveness.

4. RESULTS AND DISCUSSIONS

This study provides the performance of Decision Tree and Logistic Regression classifiers that have been developed for the classification of cybersecurity risk in e-banking services. The research includes the following stages: exploratory data analysis, selection of features, parameter setting of models, their predictive performance, analysis of the confusion matrix, precision and recall, and ROC analysis. The data consists of 104 observations and 26 independent variables. The sample was split into training and testing samples by an 80:20 rule; moreover, five-fold cross-validation was used in the process of evaluating models.

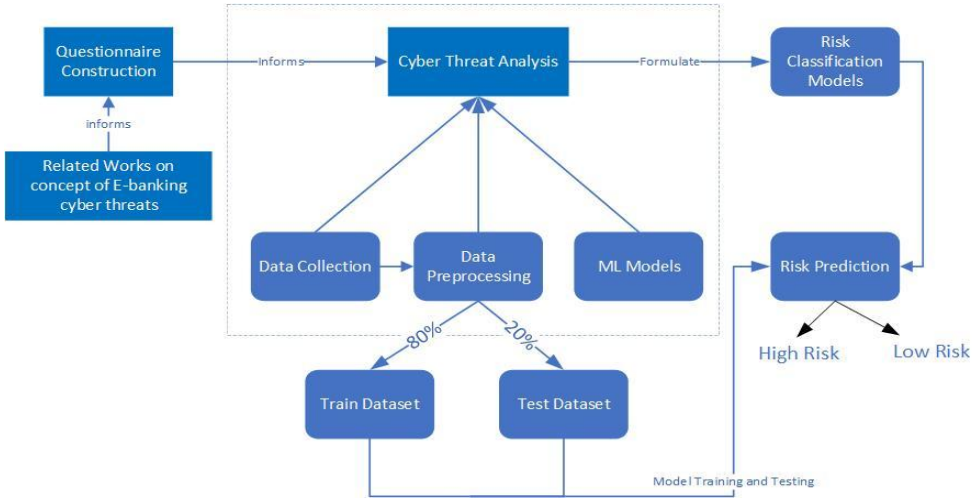

Figure 1: Methodological Diagram of the Proposed Framework

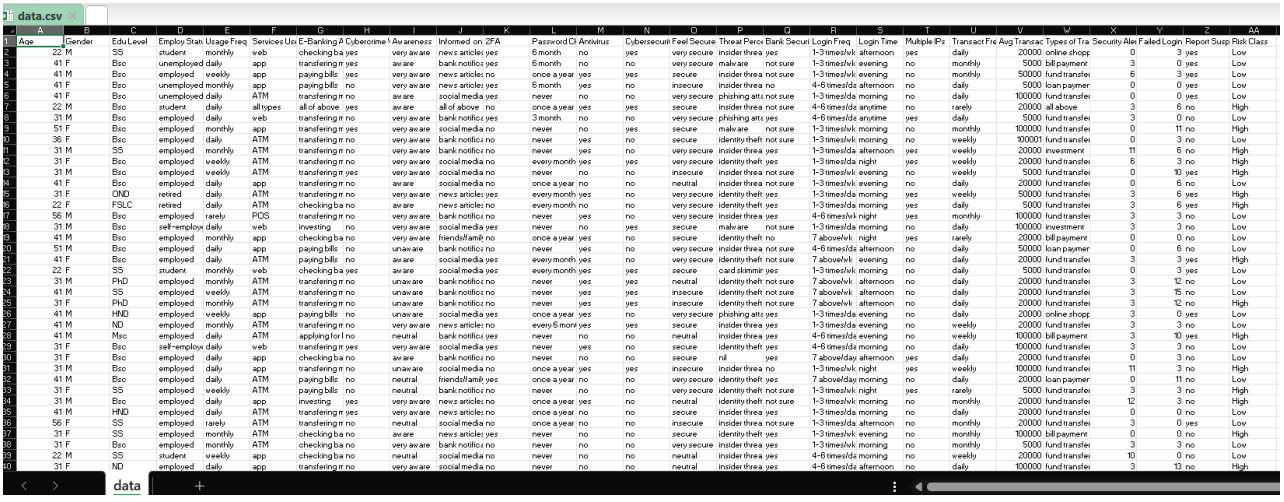

Figure 2: Cross Section of dataset

4.1.1 *Exploratory Data Analysis*

EDA allows us to understand the data better in terms of distribution. It also eliminates human assumptions about the data through the application of basic statistical tools to analyze the data. A presentation of the demographical distributions of some of the variables in the data that was gathered from the questionnaire are shown from figure 3 to 5.

Gender Distribution

The graph of figure 3 shows the bar chart of the gender distribution. From the chart, 40 observations are responds from male while the remaining were from female.

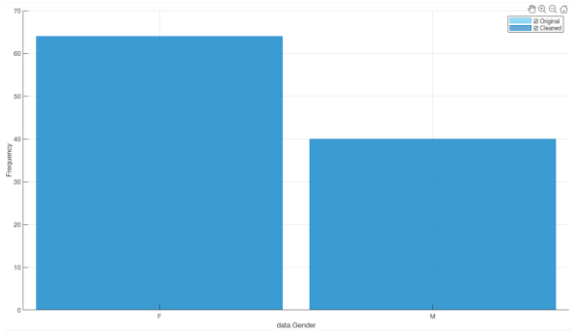

Figure 3: Gender distribution

Age and Employment Status

Figure 4 below depicts employment distribution. From figure 4 mostly respondents of age 18 to 45 response and employed users count mostly higher in the number of responses in the questionnaire.

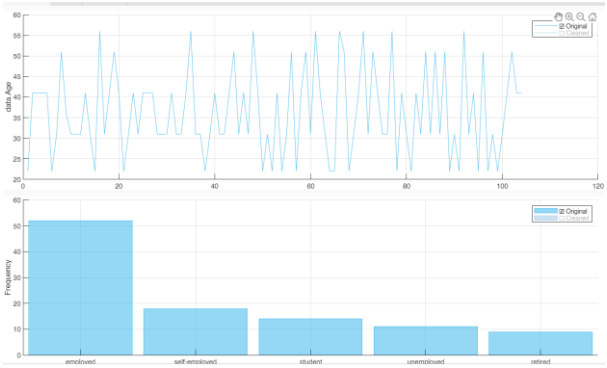

Figure 4: Age & Employment

Educational level of the respondents

The histogram plot of figure 5 shows the educational level distribution of the respondents. From the plot, a total of over 55% users where users that hold a bachelor certificate as compared to other educational level of the respondents.

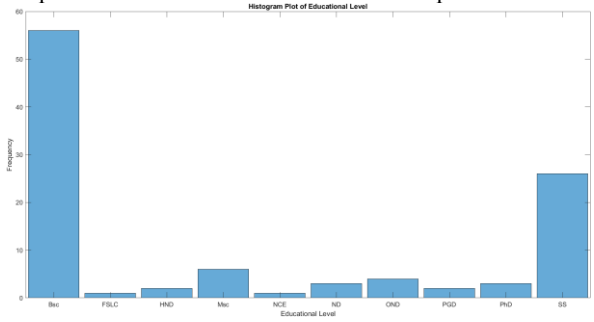

Figure 5: Educational Level Distribution

4.1.2 *Feature Selection*

Figure 6 depict the cross-section view of the feature ranked using chi-square feature importance method.

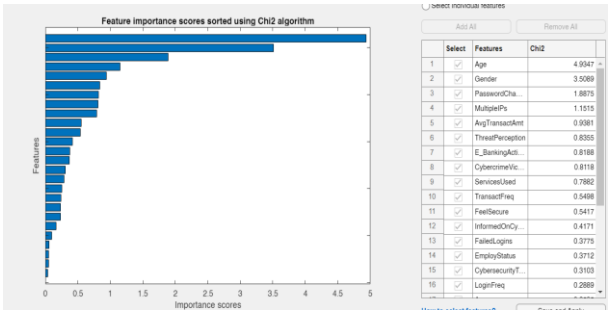

Figure 6: Chi square feature selection

4.1.3 *Hyperparameter Setting*

Table 1 summaries the model parameters setting for the Machine learning classifiers utilize in this study.

Table 1: Model Parameters Settings

Model	Parameter
Decision Tree	1. Maximum number of splits: 100 2. Split criterion: Gini's diversity index
Logistic Regression	1. Relative coefficient tolerance:0.001 2. Number of lambda value: 25

4.1.4 *Model Testing*

Table 2 shows the summary of the different classification results of the two-machine learning algorithm employed in this study. The model performances were measured based on their accuracy, precision, recall and sensitivity. Decision tree model had a 66.3% accuracy, 66.8% precision score and 66.2% score and 66% F1 score while the logistic regression classifier had 56.7% accuracy score, 56.4% precision score and 56.2% recall score and 55.9 F1 score.

Table 2: Summary of classifiers performance on test data

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Decision Tree	66.3	66.8	66.2	66.0
Logistic Regression	56.7	56.4	56.2	55.9

4.1.5 *Confusion Matrix*

The confusion matrix analysis presents an elaborate way of evaluating the classification errors for both the models. For the Logistic Regression model, 66.0% of the high-risk observations were correctly classified, while 34.0% were wrongly classified as low-risk observations. For the low-risk category on the other hand, 52.9% of the observations were

correctly classified, while 47.0% were wrongly classified as high risk. For the Decision Tree model, the classification was relatively higher in relation to the high-risk category since 75.5% of the observations were correctly classified, and only 24.5% were wrongly classified as low risk observations. In the low-risk category, 43.1% of the observations were correctly classified, while 56.9% were wrongly classified as high-risk observations. The confusion matrices for logistic regression model and decision tree model are presented below in figure 7 and 8 respectively.

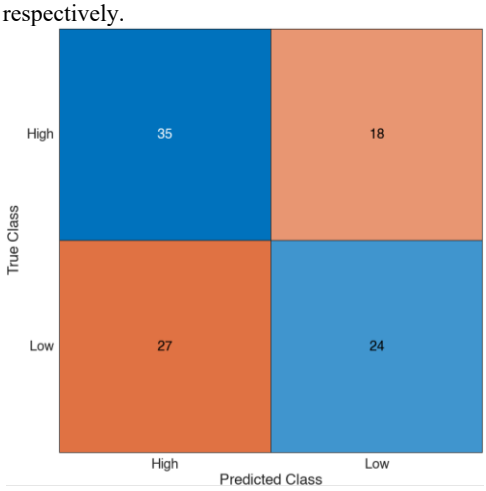

Figure 7: Confusion matrix for logistic regression model

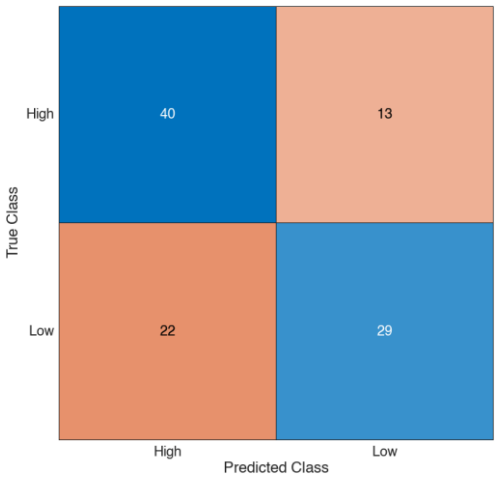

Figure 8: Confusion matrix for decision tree model on

4.1.6 Precision-Recall Curve

The precision-recall curve of the classifiers displays the positive prediction value (precision) against the true positive rate (recall) for various classification score thresholds. The accuracy and true positive rate (TPR) associated with the threshold that the classifier uses to categorize an observation are displayed by the Model Operating Point. For the Logistic Regression model, an accuracy of 0.567 indicates that approximately 56.7% of the test observations were correctly classified. The precision value of 0.564 indicates that approximately 56.4% of observations predicted as positive were actually positive. When the classifier properly allocates 57% of the positive class observations to the positive class, it has a true positive rate of 0.57. Likewise, the Decision Tree model achieved an accuracy of 66.3%, indicating that approximately 66.3% of the test observations were correctly classified as high risks are indeed high risk. The classifiers' overall quality is gauged by the precision-recall area under the curve (PR-AUC) value. Higher PR-AUC values often imply both good accuracy and high recall. The PR-AUC values range from 0 to 1. The PR-AUC for logistic regression and decision tree plots are shown in figure 9 to 10.

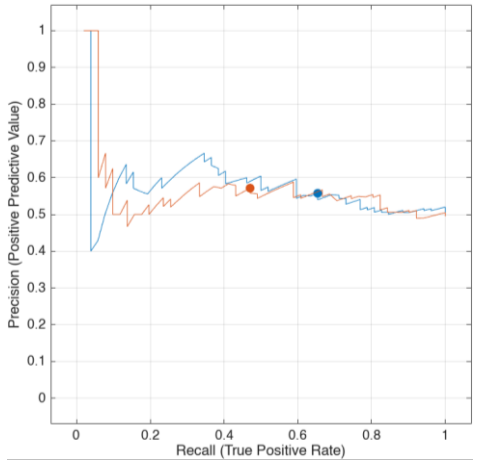

Figure 9: Precision-Recall Curve for Logistic Regression Model

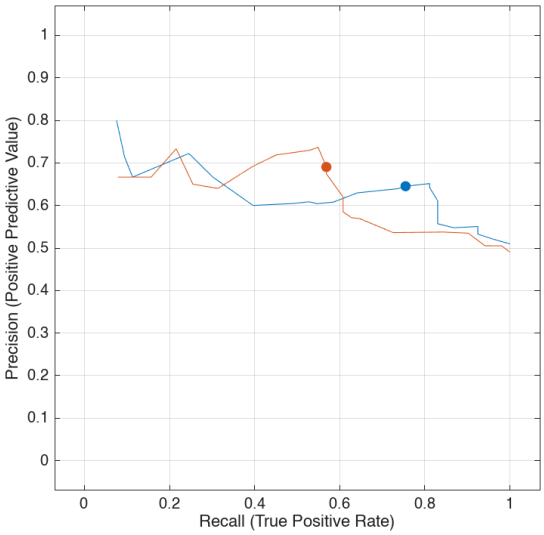

Figure 10: Precision-Recall Curve for Decision Tree Model

4.1.7 Roc-Curve

The ROC curve shows the true positive rate (TPR) versus the false positive rate (FPR) for different thresholds of classification scores, computed by the currently selected classifier. The Model Operating Point shows the false positive rate and true positive rate corresponding to the threshold used by the classifier to classify an observation. For example, a false positive rate of 0.31 indicates that the classifier incorrectly assigns 31% of the negative class observations to the positive class. A true positive rate of 0.65 indicates that the classifier correctly assigns 65% of the positive class observations to the positive class. The ROC curve for decision tree and the logistic regression model is presented in figure 11 to 12 respectively.

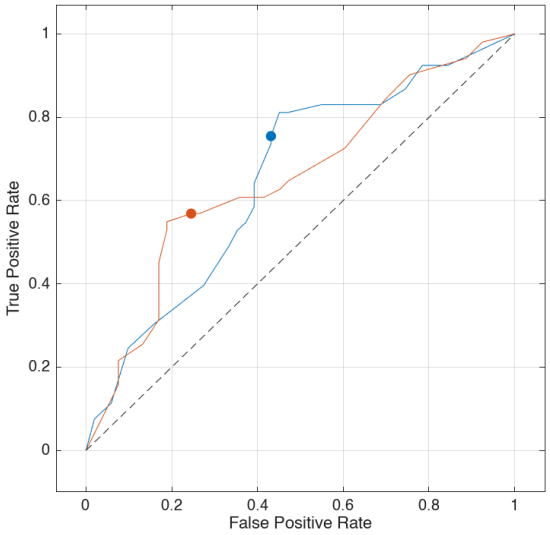

Figure 11: ROC Curve for decision tree model

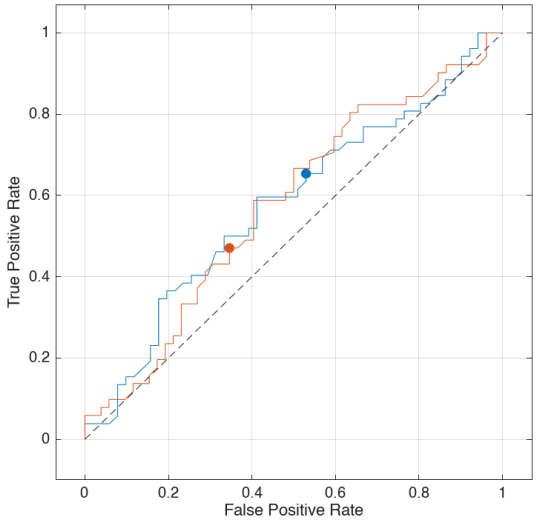

Figure 12: ROC curve for logistic regression model

4.1.8 Parallel Coordinate Curve

High dimensional data can be visualized using parallel coordinate plot. This plot allows for investigation of features that are useful for separating of classes during classification. Also, during model prediction, parallel coordinates plots identified misclassified points of each of the features. Figure 13 and 14 shows parallel coordinate plot for decision tree and logistic regression model respectively. In both figures, the dashed line in the plot denotes points that are misclassified while the correctly classified points are denoted using normal lines.

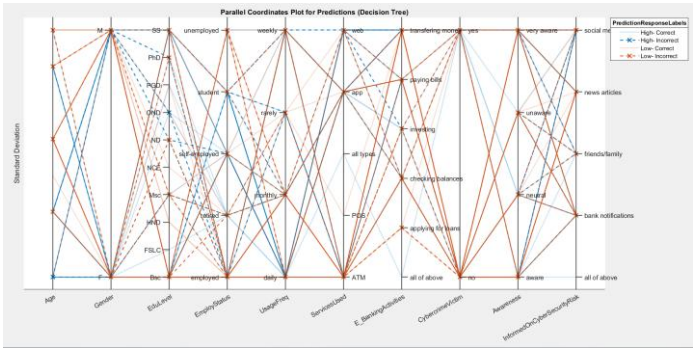

Figure 13: Parallel Coordinate Plot for Decision Tree Model Prediction

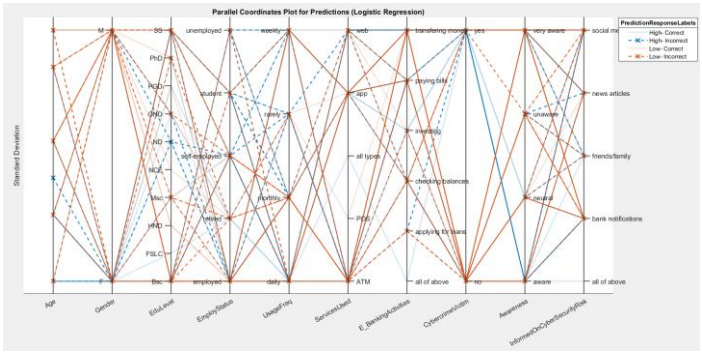

Figure 4.17: Parallel Coordinate Plot for Logistic Regression Model

5. CONCLUSION

The study successfully designed and implemented a machine learning classification model capable of effectively categorizing cybersecurity risk levels in e-banking systems, addressing a pressing need to fortify digital financial services. The research began with data collection and preprocessing, ensuring the dataset was accurate, uniform, and ready for analysis. Exploratory Data Analysis (EDA) uncovered critical patterns, relationships, and trends among variables, providing a solid foundation for data-driven decision-making. The dataset, consisting of 104 records and 26 predictive features, underwent variable transformation and normalization processes to optimize the model's scalability and performance. The chi-square feature selection technique was applied to identify the most influential variables affecting cybersecurity risk. This method reduced dimensionality while preserving essential data, allowing the model to prioritize the most critical predictors and enhance classification precision. The model was trained and tested using the test set under 5- cross fold validation. Performance metrics, including accuracy, precision, recall, and F1 score, validated the model's ability to differentiate between various levels of cybersecurity risk. This study is limited to the dataset which comprised only 104 observations collected through questionnaire responses, which may not fully capture the diversity and complexity of cybercrime risks in e-banking environments. The relatively small sample size may limit the generalizability of the developed prediction models. Further studies collected from multiple financial institutions to improve the robustness and generalizability of cybercrime risk prediction and incorporation of the explainable AI for model transparency.

6. REFERENCES

[1] B. Chaimaa, E. Najib, and H. Rachid, "E-banking Overview: Concepts, Challenges and Solutions," Wireless

Personal Communications 2020 117:2, vol. 117, no. 2, pp. 1059–1078, Nov. 2020, doi: 10.1007/S11277-020-07911-0.

[2] S. K C, "Adoption of Digital Banking in Rural Nepal : Challenges and Opportunities," 2026, Accessed: Jul. 08, 2026. [Online]. Available: http://www.theseus.fi/handle/10024/927738

[3] F. M. Alshannag, "Cyber Security Challenges in Digital Banking: Perspectives from Financial Institutions and Customers," 2026 International Conference on Artificial Intelligence for Sustainable Engineering and Innovation (AISEI), pp. 755–761, Apr. 2026, doi: 10.1109/AISEI68628.2026.11572949.

[4] B. Madhana, D. Shaik, G. Tank, A. Bahl, and M. Kanan, "An AI-driven framework for cybersecurity awareness and user trust in indian digital banking," Scientific Reports 2026, Jun. 2026, doi: 10.1038/S41598-026-55966-Z.

[5] N. Balcı, "Cybercrime in the Digital Finance Ecosystem: Typologies, Terminology, and Economic Cost," Corruption and Crime in Finance, pp. 133–154, May 2026, doi: 10.1108/978-1-83708-170-720261019.

[6] S. R. Jasim, M. Hammoodi, A. T. Hameed, and A. A. Al-Ali, "The Impact of Social Engineering on Cybersecurity Risks in the Banks," 2026 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems, ICETSIS 2026, pp. 762–766, 2026, doi: 10.1109/ICETSIS68266.2026.11548790.

[7] K. Yasmeen and M. Adnan, "Zero-day and zero-click attacks on digital banking: a comprehensive review of double trouble," Risk Management 2023 25:4, vol. 25, no. 4, pp. 25-, Sep. 2023, doi: 10.1057/S41283-023-00130-4.

[8] A. P. Ekong, A. Etuk, and M. Ekere-Obong, "Securing Against Zero-Day Attacks: A Machine Learning Approach for Classification and Organizations' Perception of its Impact," Journal of Information Systems and Informatics, vol. 5, no. 3, 2023, doi: 10.51519/journalisi.v5i3.546.

[9] N. Mohamed, "Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms," Knowledge and Information Systems 2025 67:8, vol. 67, no. 8, pp. 6969–7055, Apr. 2025, doi: 10.1007/S10115-025-02429-Y.

[10] S. Gupta, "Hacking the System: A Deep Dive into the World of E-Banking Crime," The Techno-Legal Dynamics of Cyber Crimes in Industry 5.0, pp. 121–148, Jan. 2025, doi: 10.1002/9781394242177.CH7;CTYPE:STRING:BOOK.

[11] "Cybersecurity and cybercrime: Current trends and threats," Journal of International Studies, vol. 17, no. 2, pp. 220–239, 2024.

[12] I. Riadi, Sunardi, and D. Aprilliansyah, "Analysis of Anubis Trojan Attack on Android Banking Application Using Mobile Security Labware.," International Journal of Safety & Security Engineering, vol. 13, no. 1, p. 31, Feb. 2023, doi: 10.18280/IJSSE.130104.

[13] M. Wazid, S. Zeadally, and A. K. Das, "Mobile Banking: Evolution and Threats: Malware Threats and Security Solutions," IEEE Consumer Electronics Magazine, vol. 8, no. 2, pp. 56–60, Mar. 2019, doi: 10.1109/MCE.2018.2881291.

[14] N. Chhabra Roy and S. P, "Proactive cyber fraud response: a comprehensive framework from detection to mitigation in banks," Digital Policy, Regulation and Governance, vol. 26, no. 6, pp. 678–707, Sep. 2024, doi: 10.1108/DPRG-02-2024-0029.

[15] S. Das and D. Ganguly, "Protecting Your Assets: Effective Use of Cybersecurity Measures in Banking Industries," pp. 265–286, 2024, doi: 10.1007/978-981-97-1249-6_12.

[16] M. Asmar and A. Tuqan, "Integrating machine learning for sustaining cybersecurity in digital banks," Heliyon, vol. 10, no. 17, p. e37571, Sep. 2024, doi: 10.1016/J.HELIYON.2024.E37571.

[17] lmeh Umoren, Saviour lnyang, and Onukwugha Gilean, "Bayesian Network Algorithm for Predictive Modeling of Cyber Security for Efficient Bank Channels Digitalization – Digital Library," International Journal of Information Security, Privacy on Digital Forensics, vol. 5, no. 2, Accessed: Aug. 14, 2026. [Online]. Available: https://library.ncs.org.ng/download/bayesian-network-algorithm-for-predictive-modeling-of-cyber-security-for-efficient-bank-channels-digitalization/

[18] R. Hariharan, "API Gateway Threat Prevention in Large-Scale Applications," International Journal of Sustainability and Innovation in Engineering, vol. 2, no. 1, 2024, doi: 10.56830/IJSIE202411.

[19] P. Orosz, B. Nagy, and P. Varga, "Real-Time Detection and Mitigation Strategies Newly Appearing for DDoS Profiles," Future Internet 2025, Vol. 17, vol. 17, no. 9, Aug. 2025, doi: 10.3390/FI17090400.

[20] R. Muthalagu, J. Malik, and P. M. Pawar, "Detection and prevention of evasion attacks on machine learning models," Expert Syst. Appl., vol. 266, p. 126044, Mar. 2025, doi: 10.1016/J.ESWA.2024.126044.

[21] A. N. Kalejaiye, "Adversarial Machine Learning for Robust Cybersecurity: Strengthening Deep Neural Architectures against Evasion, Poisoning, and Model-Inference Attacks," International Journal of Computer Applications Technology and Research, vol. 13, pp. 2319–8656, 2024, doi: 10.7753/IJCATR1312.1008.

[22] MohammedAnwar, "Artificial Intelligence-Powered Cyber Attacks: Adversarial Machine Learning," Feb. 2025, doi: 10.22541/AU.173862063.39098197/V1.

[23] U. K. Tripathi, S. Bagchi, S. K. Debnath, M. N. Bin Hj Mohd, A. A. Badrul Hisham, and G. Kaur, "Analyzing Cyber Threats in Banking Sector and Their Effective Solution—A Review," Lecture Notes in Networks and Systems, vol. 1593 LNNS, pp. 13–24, 2026, doi: 10.1007/978-3-032-02790-0_2/SAVE-RESEARCH.

[24] Saviour Inyang, "cybercrime concerns on e-banking services." Accessed: Jul. 09, 2026. [Online]. Available: https://www.kaggle.com/datasets/saviourjoshuainyang/cybercrime-concerns-on-e-banking-services

[25] K. Henrickson, F. Câmara Pereira, and F. Rodrigues, "Data preparation," Mobility Patterns, Big Data and Transport Analytics: Tools and Applications for

Modeling, pp. 75–116, Jan. 2026, doi: 10.1016/B978-0-443-26789-5.00008-0.

[26] S. Inyang, D. Essien, and C. E. Ndudirim, "Comparative Analysis of Machine Learning Models for Irrigation Technique Classification in Precision Agriculture," International Journal of Applied Information Systems (IJAIS, vol. 12, no. 47, 2025, Accessed: Jul. 09, 2026. [Online]. Available: www.ijais.org