Nusantara Journal of Multidisciplinary Informatics

Volume 1 Issue 1
ISSN XXXX-XXXX
https://doi.org/10.56705/njomi.v1i1.2

Research Article

Explainable Machine Learning of Scam Solicitation Exposure Among Indonesian Mobile Phone Users Using Global Findex 2025

Rahmadani[1]; Lukman Syafie[2]
[1] Universitas Muslim Indonesia, Jl. Urip Sumoharjo No.km.5, Panaikang, Kec. Panakkukang, Kota Makassar, Sulawesi Selatan 90231, Indonesia, rahmadani.iclabs@umi.ac.id
[2] Universiti Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia, lukman.syafie@s.unikl.edu.my

Correspondence should be addressed to Rahmadani; rahmadani.iclabs@umi.ac.id

Received 12 January 2026 Accepted 20 March 2026; Published 30 April 2026

Abstract:
Introduction: The expansion of mobile connectivity and digital financial services has increased financial inclusion while creating additional channels for scam solicitation. This study examines factors associated with reported scam-solicitation exposure among Indonesian mobile phone users using nationally representative Global Findex 2025 data. Method: The analysis included 850 respondents with valid scam-solicitation responses, consisting of 160 exposed and 690 non-exposed individuals. Survey weights were incorporated into descriptive analysis, model fitting, and evaluation. Logistic Regression, Random Forest, and XGBoost were assessed using stratified five-fold cross-validation. Four incremental predictor sets representing demographics, connectivity and digital capability, digital behavior, and financial participation were evaluated. SHAP was used for model interpretation. Results: The survey-weighted prevalence of reported scam-solicitation exposure was 18.54%. Random Forest achieved the strongest overall performance, with a ROC-AUC of 0.677, PR-AUC of 0.331, sensitivity of 0.590, and Brier score of 0.143. The demographic-only model achieved a ROC-AUC of 0.539, increasing to 0.617 with connectivity variables, 0.650 with digital behavior, and 0.677 with financial-participation variables. SHAP identified online activities, mobile-money ownership, digitally enabled accounts, digital payments, and online job searching among the most influential predictors. Conclusion: Digital participation provides substantially greater discriminatory information about scam-solicitation exposure than demographic characteristics alone. These findings support digital-safety interventions embedded within high-engagement digital and financial environments.

Keywords: Digital Scam; Scam Solicitation; Global Findex 2025; Explainable Machine Learning; Random Forest; SHAP; Digital Financial Inclusion; Indonesia.

1. Introduction

The rapid expansion of mobile connectivity and digital financial services has transformed how individuals communicate, access information, perform transactions, and participate in the formal financial system. Mobile phones now function not only as communication tools but also as gateways to digital payments, online commerce, employment opportunities, social media, and public services. The Global Findex Database 2025 reflects this transformation by extending internationally comparable measures beyond financial inclusion to mobile connectivity, internet use, and digital safety [1]. In Indonesia, previous analyses of Global Findex data have shown that individual characteristics such as income, education, and mobile-phone ownership are associated with access to financial accounts and digital financial services [2].

This expansion of digital participation also creates additional opportunities for fraudulent contact. Digital scams can reach potential targets through phone calls, text messages, instant messaging platforms, social media, email, and other online channels. Houtti et al. [3], using nationally representative surveys across 12 countries including Indonesia, showed that scam exposure is widespread and varies across countries, communication channels, and economic conditions. However, exposure to a scam is not equivalent to successful fraud victimization. Receiving a fraudulent solicitation represents contact with a potential offender, whereas victimization generally requires further engagement, compliance, disclosure of information, or financial loss [4].

The relationship between digital activity and scam exposure can be partly understood through routine activity theory. When applied to online environments, the theory suggests that frequent digital participation may increase the opportunity for individuals to become accessible to motivated offenders. Pratt et al. [5] found that routine online activities were associated with Internet fraud targeting, while Whitty [6] showed that both online behavior and individual characteristics influenced susceptibility to cyber-fraud. These findings indicate that digital participation is not inherently harmful, but greater involvement in online environments may create a larger opportunity surface for fraudulent contact.

Recent studies also suggest that fraud-related outcomes cannot be explained by demographic characteristics alone. Balakrishnan et al. [7] found that behavioral and psychosocial factors contributed to online fraud victimization alongside personal characteristics. In Thailand, Kraiwanit et al. [8] reported relationships between digital fraud and factors such as device use, social-media engagement, age, and income. These findings support the view that technological and behavioral exposure should be examined together with conventional demographic factors.

The Indonesian context is particularly relevant. Existing studies have documented how online fraudsters exploit familiar terminology, deceptive web addresses, and other credibility cues to mislead users [9]. Research on Indonesian online banking users has also linked previous fraud experience with fear of cybercrime and distrust in online banking [10], while Simaremare et al. [11] highlighted the role of digital literacy and cybersecurity awareness in phishing susceptibility. Beyond user-oriented cyber threats, Indonesian research has also emphasized the importance of structured digital-forensic procedures for identifying concealed or manipulated digital evidence, illustrating the broader technical challenges associated with contemporary cybercrime [12]. Although these studies confirm that fraud, phishing, and digital-security threats are important concerns in Indonesia, most focus on specific user groups, psychological factors, or particular cyber-threat mechanisms rather than nationally representative patterns of scam solicitation exposure.

Machine learning provides an additional analytical perspective for examining these patterns. Lokanan and Liu [13] demonstrated that machine-learning methods can classify fraud victimization using individual-level characteristics, while subsequent studies have incorporated explainable artificial intelligence to identify variables contributing to model predictions [14]–[16]. Indonesian machine-learning research has similarly demonstrated that algorithm performance is strongly dependent on dataset characteristics, feature representation, and validation design across different classification problems [17], [18]–[20]. Recent studies in the *Indonesian Journal of Data and Science* have further emphasized model comparison, validation rigor, and problem-specific evaluation in applications ranging from scientific-text detection and weakly supervised social-media analysis to tabular predictive modeling [21]–[23]. These findings reinforce the importance of evaluating competing algorithms empirically rather than assuming that one learning method will perform consistently across datasets.

A particularly relevant recent study applied leakage-aware explainable machine learning to fraud detection and demonstrated that apparently strong predictive performance can decline substantially after potentially leaked variables are removed [24]. Similarly, explainable machine-learning research using digital-platform behavior has shown how demographic and behavioral variables can be combined with model interpretation to characterize complex user outcomes [25]. These methodological considerations are important for the present study because scam-solicitation exposure is imbalanced, behaviorally heterogeneous, and potentially sensitive to predictor selection.

Despite this growing literature, an important gap remains in Indonesia. Earlier Global Findex studies have largely focused on account ownership, saving, borrowing, mobile money, and digital-payment adoption [2]. At the same time,

existing scam and fraud studies commonly rely on purpose-built surveys, specific user populations, transactional data, or cross-country analyses [3], [7], [8], [13]. The newly introduced connectivity and digital-safety indicators in Global Findex 2025 therefore provide an opportunity to examine scam-solicitation exposure using nationally representative Indonesian microdata while jointly considering socio-demographic characteristics, digital connectivity, online behavior, and financial participation.

This study investigates reported scam-solicitation exposure among Indonesian mobile phone users using Global Findex 2025 data. Scam solicitation is defined specifically as receiving, during the previous 12 months, a phone call, SMS, or text message from an unknown person asking the respondent to send money [1]. The study does not interpret this measure as confirmed fraud victimization.

Three research objectives are addressed. First, the study evaluates whether socio-demographic characteristics alone are sufficient to distinguish respondents reporting scam solicitation from those reporting no exposure. Second, it examines whether adding connectivity, digital behavior, and financial-participation variables improves out-of-sample discrimination using Logistic Regression, Random Forest, and XGBoost. Third, SHAP-based interpretation is used to identify which variables contribute most strongly to model predictions.

The study contributes in three ways. First, it provides Indonesia-specific evidence using nationally representative Global Findex 2025 microdata. Second, it evaluates the incremental contribution of demographic, connectivity, digital-behavior, and financial-participation domains rather than focusing only on algorithm comparison. Third, it combines survey weighting, out-of-sample machine learning, and explainability to provide population-level insight into scam-solicitation exposure without treating predictive associations as causal relationships.

2. Method

Research Design and Data

This study used a quantitative cross-sectional design combining survey-weighted analysis, machine learning, predictor-domain ablation, and explainable artificial intelligence. The data were obtained from the Indonesian microdata of the Global Findex Database 2025, identified as `IDN_2024_FINDEX_v02_M` [1], [26]. The original dataset contained 1,068 respondents and 199 variables.

Global Findex surveys nationally representative populations aged 15 years and older and provides respondent-level sampling weights to account for unequal selection probabilities and post-stratification adjustments [1]. The variable wgt was therefore incorporated in descriptive analysis, model fitting, and performance evaluation.

The outcome was derived from con21, which asks whether respondents had received, during the previous 12 months, a phone call, SMS, or text message from an unknown person asking them to send money. Responses were coded as 1 for reported scam-solicitation exposure and 0 for no reported exposure. "Don't know" and invalid responses were excluded.

Among the 1,068 respondents, 851 had responses to con21. One "Don't know" response was removed, resulting in a final analytic sample of 850 respondents, consisting of 160 exposed and 690 non-exposed individuals. The outcome represents exposure to scam solicitation rather than confirmed fraud victimization.

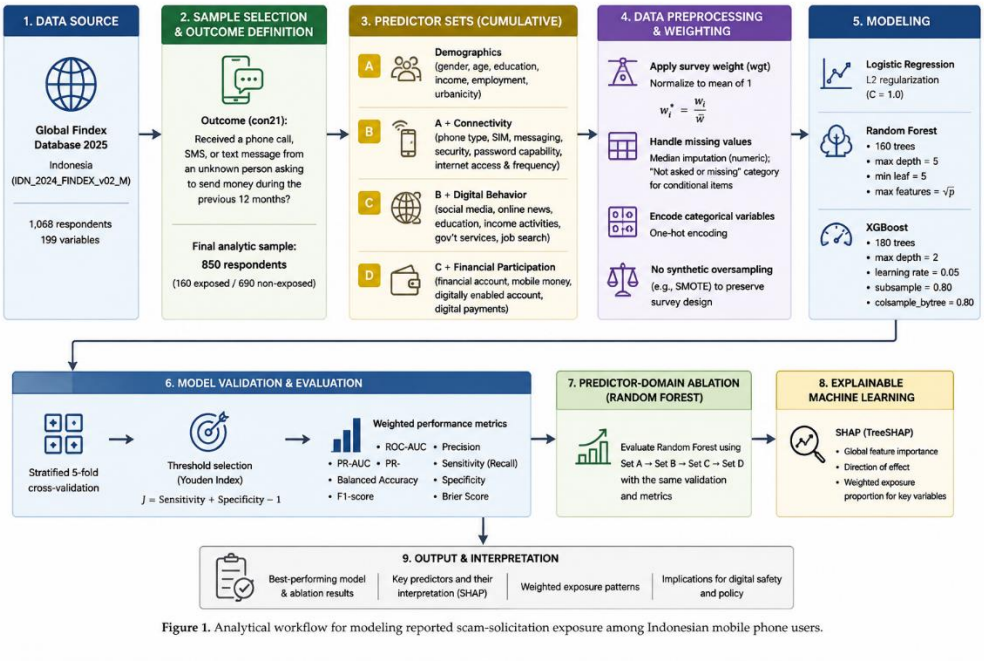

Figure 1. Analytical workflow for modeling reported scam-solicitation exposure among Indonesian mobile phone users.

Predictor Variables and Preprocessing

Predictors were organized into four cumulative domains to evaluate whether information beyond basic demographics improved model discrimination.

Table 1. Predictor sets used in the analysis

Set	Predictor domain	Main variables
A	Demographics	Gender, age, education, income quintile, employment, urbanicity
B	Set A + connectivity	Phone type, SIM-related characteristics, messaging capability, phone security, password capability, internet access and frequency
C	Set B + digital behavior	Social media, online news, education, online income generation, government services, and online job search
D	Set C + financial participation	Financial account, mobile money, digitally enabled account, and digital-payment activity

Variables directly related to the outcome, including `con22`, were excluded to prevent target leakage. This decision was particularly important because recent explainable fraud-detection research has demonstrated that post-outcome or decision-related variables can substantially inflate apparent predictive performance [25].

Age was treated as a continuous variable, while the remaining predictors were modeled as categorical variables. Missing numerical values were imputed using the training-set median. Categorical variables were transformed using one-hot encoding. Responses indicating "Don't know" or "Refused" were treated as missing.

Several Global Findex questions were conditionally administered based on earlier responses. These structural missing values were retained as a separate "Not asked or missing" category rather than being automatically interpreted as negative responses.

No synthetic oversampling method such as SMOTE was applied because the analysis used survey microdata and population weights. This avoided introducing artificial observations that could distort the original survey structure.

Survey Weighting

The sampling weight wgt was normalized to a mean of one before model fitting:

$$w_i^* = \frac{\mathrm{w_i}}{\overline{w}} \tag{1}$$

where (w_i) is the original survey weight and ($\overline{w}$) is its mean.

Normalization preserved the relative contribution of each respondent while improving numerical stability during model training. Survey weights were also applied when calculating performance metrics on held-out observations.

For descriptive purposes, the effective sample size was calculated using the Kish formulation:

$$n_{eff} = \frac{(\sum_{i=1}^{n} w_i)^2}{\sum_{i=1}^{n} w_i^2} \tag{2}$$

which produced an effective sample size of approximately 664.9 respondents.

Machine-Learning Models

Three classification algorithms were evaluated: Logistic Regression, Random Forest, and XGBoost. Logistic Regression was included as a linear and relatively interpretable baseline. The model used L2 regularization with (C=1.0). Random Forest was selected to capture nonlinear relationships and interactions among heterogeneous survey variables [27]. The model used 160 trees, maximum depth of 5, minimum leaf size of 5, and square-root feature sampling. XGBoost was included as a regularized boosting approach [28]. The configuration used 180 trees, maximum depth of 2, learning rate of 0.05, row subsampling of 0.80, and feature subsampling of 0.80.

Fixed and deliberately restrained model configurations were used instead of extensive hyperparameter optimization. This reduced the risk of overfitting model selection to the relatively modest analytic sample and improved reproducibility.

The use of multiple candidate models and out-of-sample validation was also supported by previous Indonesian machine-learning studies. Azis and collaborators have reported varying performance across classification architectures, feature representations, and datasets [18], [19], [20], while Random Forest combined with five-fold cross-validation has also been evaluated in applied prediction tasks [21]. Recent work in IJODAS similarly demonstrates the importance of rigorous validation when comparing predictive models [23]. These studies support the present decision to evaluate the algorithms empirically rather than selecting a model solely from prior performance reported in another application domain.

Model Validation and Evaluation

Performance was assessed using stratified five-fold cross-validation. The class distribution was approximately preserved within each fold. In every iteration, four folds were used for training and the remaining fold for independent evaluation. All three algorithms used identical fold assignments.

The classification threshold was determined from the training data using the Youden index [29]:

$$J = Sensitivity + Specificity - 1 \tag{3}$$

The selected threshold was then applied unchanged to the corresponding test fold, ensuring that test observations did not influence threshold selection.

Because the exposed class was less frequent, evaluation was not based on accuracy alone. The following metrics were calculated using survey weights:

- ROC-AUC;
- PR-AUC;
- balanced accuracy;

- precision;
- sensitivity;
- specificity;
- F1-score; and
- Brier score.

PR-AUC was included because precision-recall analysis is particularly useful for imbalanced classification problems [30]. The Brier score was used to assess the quality of predicted probabilities [31]. Model results were summarized as the mean and standard deviation across the five test folds.

Predictor-Domain Ablation

After comparing the three algorithms using the complete predictor set, the model with the highest mean weighted ROC-AUC was selected for additional analysis. Random Forest achieved the strongest overall discrimination and was therefore evaluated sequentially using predictor Sets A through D. This ablation procedure examined whether adding connectivity, digital behavior, and financial-participation variables improved discrimination beyond demographic information alone. The same five-fold validation procedure, survey weighting, threshold selection, and evaluation metrics were applied to every predictor set.

Explainable Machine Learning

The final Random Forest model was interpreted using SHapley Additive exPlanations (SHAP) [32]. TreeSHAP was applied to estimate the contribution of individual predictors to model predictions. Because categorical survey variables generated multiple one-hot encoded features, absolute SHAP values belonging to the same original variable were aggregated to obtain variable-level importance:

$$I_j = \sum_{k \in G_j} mean(|\emptyset_k|) \tag{4}$$

where (G_j) represents the encoded features originating from variable (j). Survey-weighted exposure proportions were additionally calculated for selected high-ranking predictors to assist interpretation of the SHAP results. These comparisons were descriptive and were not treated as adjusted or causal effects. SHAP values were interpreted only as indicators of how strongly a variable contributed to the fitted model. They were not interpreted as evidence that the predictor caused scam-solicitation exposure.

Software and Reproducibility

The analysis was conducted in Python using pandas and NumPy for data processing, scikit-learn for preprocessing, cross-validation and Logistic Regression/Random Forest, XGBoost for gradient boosting, and SHAP for explainability. A fixed random seed of 42 was used throughout the analysis.

The complete analytical pipeline was implemented programmatically from raw microdata through preprocessing, model validation, ablation analysis, and SHAP interpretation. Only analytical code and aggregated outputs were retained for reproducibility; the original World Bank microdata were not redistributed.

3. Result and Discussion

Scam-Solicitation Exposure

The final analytic sample consisted of 850 Indonesian mobile phone users with valid responses to con21. Among them, 160 respondents reported receiving a phone call, SMS, or text message from an unknown person asking them to send money, while 690 reported no such exposure. The raw exposure proportion was 18.82%, and the survey-weighted estimate was 18.54%.

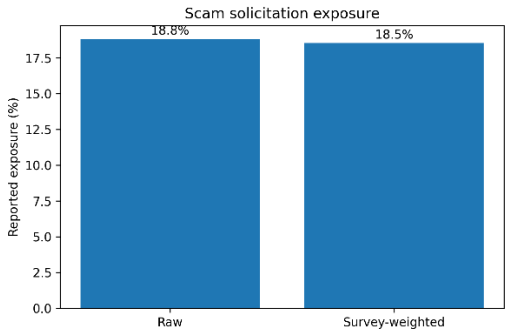

Figure 2. Raw and survey-weighted prevalence of reported scam-solicitation exposure.

This result indicates that scam solicitation through mobile communication represents a meaningful digital-safety issue. However, the outcome should be interpreted as exposure, not confirmed fraud victimization. Respondents who received scam contact did not necessarily interact with the sender or experience financial loss. Similar distinctions between scam exposure and victimization have been emphasized in previous cross-country research [3].

Model Performance

Table 2 compares the three classification algorithms using the complete predictor set.

Table 2. Survey-weighted model performance

Model	ROC-AUC	PR-AUC	Balanced Accuracy	Sensitivity	Specificity	F1-score	Brier Score
Random Forest	**0.677 ± 0.024**	**0.331 ± 0.029**	0.619	**0.59**	0.648	0.374	**0.143**
XGBoost	0.666 ± 0.039	0.323 ± 0.053	**0.619**	0.554	**0.684**	**0.378**	0.148
Logistic Regression	0.642 ± 0.036	0.316 ± 0.066	0.573	0.46	0.685	0.325	0.151

Random Forest achieved the highest ROC-AUC and PR-AUC and the lowest Brier score, indicating the strongest overall discrimination and probability quality. XGBoost produced slightly higher specificity and F1-score, while Logistic Regression showed the weakest overall performance.

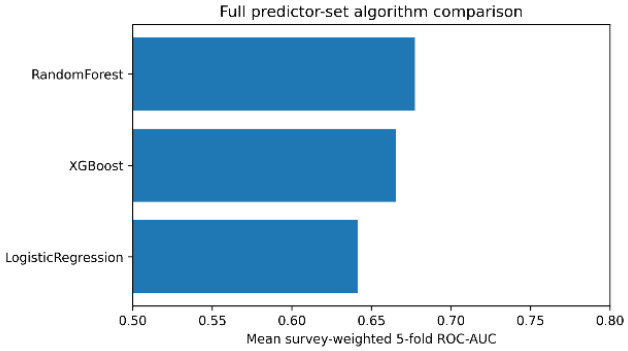

Figure 3. Survey-weighted performance comparison of Logistic Regression, Random Forest, and XGBoost.

The ROC-AUC of 0.677 indicates moderate rather than high discrimination. This is reasonable because scam exposure is influenced by factors that are not captured in Global Findex, such as telephone-number leakage, scam campaigns, communication platforms, offender strategies, and timing. The purpose of the model is therefore not to predict with certainty who will receive a scam message, but to identify population-level characteristics associated with differences in reported exposure.

The variation observed across algorithms is consistent with previous comparative machine-learning research showing that model superiority is highly dependent on the structure of the underlying data. Studies involving medical images and transfer-learning representations have reported substantial changes in performance across architectures and classifier combinations [18], [19], [20]. Similar evidence is found in recent IJODAS research, where conventional machine learning, transformer-based models, and rigorously validated ensemble approaches produced different strengths depending on the prediction problem and data-generation process [22]–[24]. Therefore, the relatively small difference between Random Forest and XGBoost in the present study is more appropriately interpreted as evidence of dataset-specific model behavior rather than universal superiority of Random Forest.

The better performance of nonlinear tree-based models suggests that relationships between digital behavior and scam exposure may involve interactions or nonlinear patterns that are not fully captured by Logistic Regression. Previous fraud research has similarly shown that machine-learning approaches can provide useful discrimination when multiple individual and behavioral characteristics are considered jointly [13].

Contribution of Digital Context

The predictor-domain ablation analysis produced the clearest result of the study.

Table 3. Random Forest performance across predictor domains

Predictor Set	Information Included	ROC-AUC	PR-AUC	F1-score	Brier Score
Set A	Demographics	0.539	0.214	0.26	0.153
Set B	+ Connectivity	0.617	0.262	0.35	0.148
Set C	+ Digital behavior	0.65	0.323	0.347	0.144
Set D	+ Financial participation	**0.677**	**0.331**	**0.374**	**0.143**

Demographic variables alone produced a ROC-AUC of only 0.539, indicating very limited discriminatory value. Adding connectivity and digital-capability variables increased ROC-AUC to 0.617. The inclusion of digital behavior further increased it to 0.650, while the full model including financial participation reached 0.677.

Overall, ROC-AUC increased by approximately 0.138 from the demographic-only model to the complete model. PR-AUC also increased from 0.214 to 0.331.

These results indicate that digital participation provides substantially more information about scam-solicitation exposure than demographic characteristics alone. Gender, age, education, income, employment, and urbanicity were insufficient to meaningfully distinguish exposed respondents. Instead, information describing connectivity, online activity, and digital-financial participation provided the largest improvement.

This pattern is consistent with routine activity explanations of online fraud, which emphasize that frequent participation in digital environments can increase opportunities for contact with potential offenders [5], [6]. Recent studies have also shown that fraud-related outcomes are associated not only with demographic characteristics but also with technology use and online behavior [7], [8].

SHAP Interpretation

SHAP analysis was used to identify the predictors that contributed most strongly to the Random Forest model.

Table 4. Top predictors according to aggregated SHAP importance

Rank	Predictor	Mean |SHAP|
1	Read news/current events online	**0.021**
2	Earned money online	**0.017**
3	Mobile money account	**0.0141**
4	Digitally enabled account	**0.0131**

Rank	Predictor	Mean |SHAP|
5	Any digital payment	**0.0119**
6	Employment status	**0.0113**
7	Any financial account	**0.0095**
8	Searched/applied for jobs online	**0.0088**
9	Ability to change phone PIN/password	**0.0087**
10	Sent voice messages using a mobile phone	**0.0083**

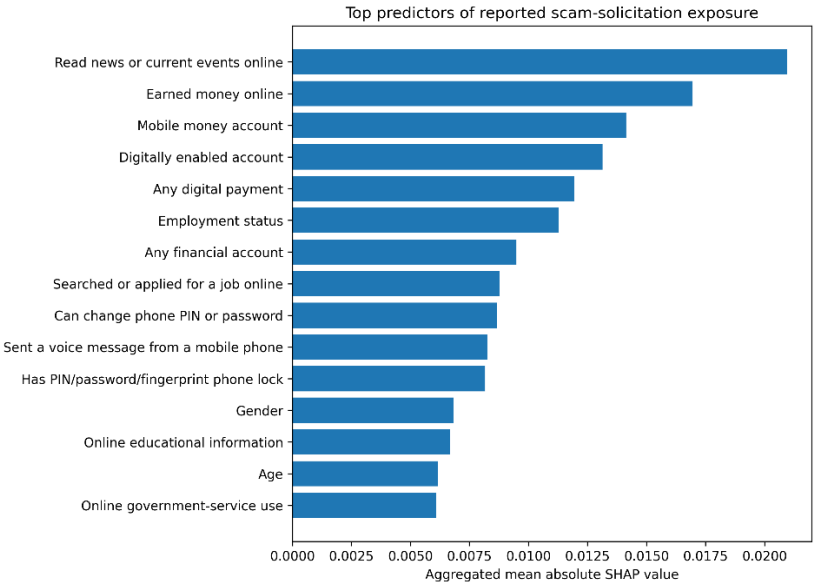

Figure 4. Aggregated SHAP variable importance for the final Random Forest model.

The highest-ranking predictors were dominated by online activity and financial participation. Survey-weighted descriptive comparisons supported this pattern. Respondents who reported earning money online showed an exposure rate of 37.42%, compared with 15.90% among those who did not. Mobile-money account holders showed exposure of 29.43%, compared with 14.48% among non-holders. Similarly, exposure was 25.05% among digital-payment users, compared with 12.18% among non-users.

Respondents who searched or applied for jobs online also showed higher reported exposure, at 31.41%, compared with 17.15% among those who did not. These findings suggest that individuals who are more active in digital environments may have more opportunities to encounter scam solicitations. The importance of behavioral variables is also consistent with recent explainable machine-learning applications in which patterns of digital-platform usage contributed substantial information beyond basic demographics [33]. At the same time, the leakage-aware fraud-detection study of Zulfikar et al. [25] demonstrates why high feature importance must be interpreted in relation to the data-generation process and predictor availability. In the present study, downstream scam-response information was deliberately excluded so that the explanation reflects characteristics available independently of confirmed victimization behavior.

However, the findings should not be interpreted causally. Using mobile money, digital payments, online news, or online employment services does not necessarily increase scam risk directly. These activities may simply indicate greater digital engagement and a larger exposure surface. Likewise, security-related variables such as PIN or password capability may reflect higher digital activity rather than vulnerability.

SHAP values therefore indicate how strongly a feature contributed to model predictions, not whether the feature caused the observed outcome. This distinction is particularly important in fraud research, where explainability methods can easily be overinterpreted [17].

Practical Implications and Limitations

The findings suggest that scam-prevention strategies should not rely only on demographic profiling. The weak performance of the demographic-only model indicates that age, gender, income, or education alone may provide limited information for identifying groups more likely to report scam contact.

Instead, preventive interventions may be more useful when integrated into high-engagement digital environments, including mobile-money services, digital-payment platforms, online job services, and other online activities. Financial institutions, telecommunications providers, online platforms, and public agencies could use these digital touchpoints to provide scam warnings, verification guidance, and reporting mechanisms.

The moderate model performance also means that the framework is more suitable for population-level analysis and policy insight than individual risk scoring. It should not be used to label particular users as likely scam targets.

Several limitations remain. First, the Global Findex data are cross-sectional, so causal relationships cannot be established. Second, con21 measures a specific form of scam solicitation and does not represent all types of phishing, identity theft, investment fraud, or online deception. Third, the positive class consisted of 160 respondents, limiting the amount of information available for model learning. Fourth, SHAP values reflect the fitted model rather than causal effects. Finally, the study relied on internal cross-validation rather than an independent temporal or external dataset.

Future studies could compare scam exposure across countries or future Global Findex waves and distinguish more explicitly between exposure, engagement, victimization, and financial loss.

4. Conclusion

This study examined reported scam-solicitation exposure among Indonesian mobile phone users using nationally representative Global Findex 2025 microdata and a survey-weighted explainable machine-learning framework. Among 850 respondents with valid outcome information, 160 reported receiving a phone call, SMS, or text message from an unknown person asking them to send money. The survey-weighted exposure proportion was 18.54%, indicating that this form of scam contact represents a meaningful digital-safety concern among Indonesian mobile phone users.

Among the evaluated models, Random Forest achieved the strongest overall discrimination, with a mean ROC-AUC of 0.677 and PR-AUC of 0.331, while XGBoost produced slightly higher specificity and F1-score. The moderate predictive performance indicates that scam-solicitation exposure cannot be explained completely by individual-level survey characteristics. Nevertheless, the predictor-domain ablation analysis revealed a clear pattern. A model using demographic characteristics alone achieved a ROC-AUC of only 0.539, while the addition of connectivity, digital behavior, and financial-participation variables progressively increased ROC-AUC to 0.677. This suggests that the way individuals participate in digital environments provides substantially more discriminatory information than demographic characteristics alone.

The explainability analysis supported this finding. Online information seeking, online income-generating activities, mobile-money ownership, digitally enabled accounts, digital payments, employment-related online activities, and other forms of digital participation were among the most influential predictors in the Random Forest model. These findings should not be interpreted as evidence that digital financial services or online activities cause scam exposure. Rather, they indicate that greater participation in the digital ecosystem may correspond to a larger opportunity surface through which fraudulent actors can initiate contact.

From a practical perspective, the results suggest that scam-prevention strategies should not rely exclusively on demographic profiling. Consumer-protection measures may be more effective when integrated into digital environments where users are actively conducting financial transactions, searching for employment, accessing information, or engaging with other online services. Financial institutions, telecommunications providers, digital platforms, and public agencies can use such insights to design more context-sensitive awareness, verification, and reporting mechanisms.

The study is limited by its cross-sectional design, self-reported outcome, modest number of positive cases, and the absence of external temporal validation. SHAP values also describe predictive contributions rather than causal effects. Future studies could extend the analysis across multiple countries or future Global Findex waves, evaluate temporal stability, and distinguish more explicitly between scam exposure, engagement, successful deception, and financial loss. Despite these limitations, the study demonstrates that combining nationally representative survey data, survey weighting, predictor-domain ablation, and explainable machine learning can provide a useful framework for understanding digital scam exposure in an increasingly connected society.

References

[1] L. Klapper, D. Singer, L. Starita, and A. Norris, The Global Findex Database 2025: Connectivity and Financial Inclusion in the Digital Economy. Washington, DC, USA: World Bank, 2025, doi: https://doi.org/10.1596/978-1-4648-2204-9.

[2] P. Prasetyoputra, Y. Isnasari, A. P. S. Prasojo, and I. Hermawan, "Factors associated with financial inclusion in Indonesia before and during COVID-19: Evidence from Global Findex data," Statistics, Optimization & Information Computing, vol. 15, no. 1, pp. 93–127, 2026, doi: https://doi.org/10.19139/soic-2310-5070-2852.

[3] M. Houtti, A. Roy, V. N. R. Gangula, and A. M. Walker, "A survey of scam exposure, victimization, types, vectors, and reporting in 12 countries," Journal of Online Trust and Safety, vol. 2, no. 4, 2024, doi: https://doi.org/10.54501/jots.v2i4.204.

[4] G. Norris, A. Brookes, and D. Dowell, "The psychology of internet fraud victimisation: A systematic review," Journal of Police and Criminal Psychology, vol. 34, no. 3, pp. 231–245, 2019, doi: https://doi.org/10.1007/s11896-019-09334-5.

[5] T. C. Pratt, K. Holtfreter, and M. D. Reisig, "Routine online activity and internet fraud targeting: Extending the generality of routine activity theory," Journal of Research in Crime and Delinquency, vol. 47, no. 3, pp. 267–296, 2010, doi: https://doi.org/10.1177/0022427810365903.

[6] M. T. Whitty, "Predicting susceptibility to cyber-fraud victimhood," Journal of Financial Crime, vol. 26, no. 1, pp. 277–292, 2019, doi: https://doi.org/10.1108/JFC-10-2017-0095.

[7] V. Balakrishnan, U. Ahhmed, and F. Basheer, "Personal, environmental and behavioral predictors associated with online fraud victimization among adults," PLOS ONE, vol. 20, no. 1, Art. no. e0317232, 2025, doi: https://doi.org/10.1371/journal.pone.0317232.

[8] T. Kraiwanit, P. Limna, R. Sonsuphap, and V. Chutipat, "Predictors of digital fraud: Evidence from Thailand," Journal of Risk and Financial Management, vol. 18, no. 12, Art. no. 671, 2025, doi: https://doi.org/10.3390/jrfm18120671.

[9] A. Ardoni, "Information items used by online fraudsters and its relationship to Indonesian digital literacy," Berkala Ilmu Perpustakaan dan Informasi, vol. 18, no. 2, pp. 326–337, 2022, doi: https://doi.org/10.22146/bip.v18i2.5866.

[10] S. Lestari, W. R. Adawiyah, A. L. Alhamidi, J. Prayogi, and R. Haryanto, "Navigating perilous seas: Unmasking online banking frauds, perceived usefulness, fear of cybercrime and distrust in online banking," Safer Communities, vol. 23, no. 4, pp. 444–464, 2024, doi: https://doi.org/10.1108/SC-04-2024-0018.

[11] H. Simaremare, M. Fikri, and Z. Shukur, "Confirmatory factor analysis of phishing susceptibility in Indonesia," International Journal of Advanced Computer Science and Applications, vol. 17, no. 6, 2026, doi: https://doi.org/10.14569/IJACSA.2026.0170638.

[12] A. Asran, E. I. Alwi, and H. Azis, "Implementasi metode static forensics untuk ekstraksi file steganografi pada bukti digital menggunakan framework NIST," JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), vol. 11, no. 1, pp. 544–556, 2026, doi: https://doi.org/10.29100/jipi.v11i1.7406.

[13] M. Lokanan and S. Liu, “Predicting fraud victimization using classical machine learning,” Entropy, vol. 23, no. 3, Art. no. 300, 2021, doi: https://doi.org/10.3390/e23030300.

[14] H. Wang and D. Wang, “Constructing online fraud victimization models: A machine learning approach,” Journal of Forensic Psychology Research and Practice, early access, 2026, doi: https://doi.org/10.1080/24732850.2026.2643611.

[15] G. S. Erbuğa and C. Ünal, “Explainable artificial intelligence-driven fraud detection: Behavioral proxies of digital financial literacy in online payment models,” Borsa Istanbul Review, Art. no. 100840, 2026, doi: https://doi.org/10.1016/j.bir.2026.100840.

[16] Y. Zhou, H. Li, Z. Xiao, and J. Qiu, “A user-centered explainable artificial intelligence approach for financial fraud detection,” Finance Research Letters, vol. 58, Part A, Art. no. 104309, 2023, doi: https://doi.org/10.1016/j.frl.2023.104309.

[17] U. Zafar and F. Wu, “Methodological challenges in explainable AI for fraud detection: A systematic literature review,” Artificial Intelligence Review, vol. 59, Art. no. 115, 2026, doi: https://doi.org/10.1007/s10462-026-11516-7.

[18] H. Azis, Y. Salim, and Y. Puspitasari, “Deep learning approaches for soft tissue classification of cephalometric images in orthodontics,” *JOIV: International Journal on Informatics Visualization*, vol. 10, no. 3, pp. 988–995, 2026, doi: https://doi.org/10.62527/joiv.10.3.3885.

[19] H. Azis, R. A. Jalil, and A. R. Manga’, “Comparative performance of VGG16 and EfficientNetB0-based transfer learning for brain tumor classification,” Knowledge Engineering and Data Science, vol. 8, no. 2, pp. 185–197, 2025, doi: https://doi.org/10.17977/um018v8i22025p185-197.

[20] A. R. Manga, A. P. Utami, H. Azis, Y. Salim, and A. Faradibah, “Optimizing classification models for medical image diagnosis: A comparative analysis on multi-class datasets,” Computer Science and Information Technologies, vol. 5, no. 3, pp. 205–214, 2024, doi: https://doi.org/10.11591/csit.v5i3.p205-214.

[21] H. Azis and N. Rismayanti, “Prediksi anemia dari pixel gambar dan level hemoglobin menggunakan Random Forest Classifier,” NERO (Networking Engineering Research Operation), vol. 9, no. 1, pp. 21–34, 2024, doi: https://doi.org/10.21107/nero.v9i1.27916.

[22] Aldo, A. W. Murdiyanto, and U. S. Aesyi, “Zero-shot detection of IndoT5-synthesized Indonesian scientific abstracts using mDeBERTa v3,” Indonesian Journal of Data and Science, vol. 7, no. 2, pp. 333–349, 2026, doi: https://doi.org/10.56705/ijodas.v7i2.457.

[23] E. H. Saputra, I. A. E. Zaeni, D. D. Prasetya, A. M. Zain, W. Antonius, and I. M. Wirawan, “Comparative evaluation of machine learning models for heavy crude oil viscosity prediction using repeated nested cross-validation and independent holdout testing,” Indonesian Journal of Data and Science, vol. 7, no. 2, 2026, doi: https://doi.org/10.56705/ijodas.v7i2.455.

[24] M. R. Hardika and B. Miftahurrohmah, “Weakly supervised sentiment analysis of gold price discussions using conventional machine learning and IndoBERT,” Indonesian Journal of Data and Science, vol. 7, no. 2, pp. 274–290, 2026, doi: https://doi.org/10.56705/ijodas.v7i2.445.

[25] D. H. Zulfikar, E. S. J. Atmadji, and B. S. W. Poetro, “Leakage-aware and explainable machine learning for healthcare claim fraud detection using imbalanced medical insurance data,” International Journal of Artificial Intelligence in Medical Issues, vol. 4, no. 1, pp. 19–34, 2026, doi: https://doi.org/10.56705/z3207345.

[26] Development Research Group, Finance and Private Sector Development Unit, Indonesia—The Global Findex Database 2025: Connectivity and Financial Inclusion in the Digital Economy, Ref. IDN_2024_FINDEX_v02_M, World Bank, 2025, doi: https://doi.org/10.48529/CDK5-2M94.

[27] L. Breiman, “Random forests,” Machine Learning, vol. 45, pp. 5–32, 2001, doi: https://doi.org/10.1023/A:1010933404324.

[28] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD ’16), San Francisco, CA, USA, 2016, pp. 785–794, doi: https://doi.org/10.1145/2939672.2939785.

[29] W. J. Youden, “Index for rating diagnostic tests,” Cancer, vol. 3, no. 1, pp. 32–35, 1950, doi: https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.

[30] G. W. Brier, “Verification of forecasts expressed in terms of probability,” Monthly Weather Review, vol. 78, no. 1, pp. 1–3, 1950, doi: https://doi.org/10.1175/1520-0493(1950)078<0001>2.0.CO;2.

[31] T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,” PLOS ONE, vol. 10, no. 3, Art. no. e0118432, 2015, doi: https://doi.org/10.1371/journal.pone.0118432.

[32] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in Advances in Neural Information Processing Systems 30, 2017, pp. 4765–4774.

[33] A. Halid, D. A. Purnamasari, A. C. Saputra, and N. M. Setiohardjo, “Explainable machine learning for predicting the mental health impact of AI and digital platform usage among students,” International Journal of Artificial Intelligence in Medical Issues, vol. 4, no. 1, pp. 1–18, 2026, doi: https://doi.org/10.56705/pxn6qg39.