MAIN ORGANISER:

OFFICIAL COLLABORATORS:

THE INTERNATIONAL COMPETITION ON SUSTAINABLE EDUCATION

20TH AUGUST 2025

TRANSFORMING EDUCATION, DRIVING INNOVATION AND ADVANCING LIFELONG LEARNING FOR EMPOWERED WORLD

SUSED2025
THE FOURTH INTERNATIONAL COMPETITION ON SUSTAINABLE EDUCATION 2025
TRANSFORMING EDUCATION, DRIVING INNOVATION AND ADVANCING LIFELONG LEARNING FOR EMPOWERED WORLD
e-PROCEEDING
20TH AUGUST 2025
ACADEMY OF LANGUAGE STUDIES
UNIVERSITI TEKNOLOGI MARA
NEGERI SEMBILAN BRANCH
KUALA PILAH CAMPUS
MALAYSIA

PREDICTING FACTORS IN FINANCIAL LOSS AMONG MALAYSIAN SCAM VICTIMS USING MACHINE LEARNING

Nur Alisa Binti Azian* & Che Norhalila Binti Che Mohamed

Faculty of Computer and Mathematical Sciences, UiTM Negeri Sembilan, Seremban Campus*

nrlszn34@gmail.com*

ABSTRACT

This study presents an innovative educational approach to scam prevention by using logistic regression and decision tree models to identify key predictors of financial loss among 394 Malaysian scam victims. Emotional harm, age, and cybersecurity knowledge emerged as the most significant factors, with emotional harm being the strongest predictor of these factors. The decision tree model demonstrated superior accuracy and interpretability compared to logistic regression, making it a practical tool for educational use. By integrating data science with digital literacy, this research supports the development of targeted learning modules and public awareness strategies. The findings emphasize the use of machine learning to enhance risk education, empower self-assessment, and inform evidence-based interventions aimed at reducing scam victimization in Malaysia.

Keywords: *Decision Tree, Financial Loss, Logistic Regression, Machine Learning Models, Malaysian Scam Victims*

INTRODUCTION

The widespread adoption of the internet has significantly transformed communication, commerce, and daily routines, making it one of the most impactful innovations in modern history (Brooks, 2022). In Malaysia, internet usage continues to rise, increasing from 96.8% in 2021 to 97.4% in 2022 (DOSM). However, this growing digital connectivity has also increased exposure to cyber threats, particularly online scams. Although scams have existed long before the internet, their tactics have evolved to exploit psychological, social, and technological vulnerabilities (Hanoch & Wood, 2021; Whitty, 2020). In Malaysia, common scam types include love scams, Macau scams, parcel scams, job scams, and online loan scams. Between 2021 and April 2024, more than 95,000 online scam cases were reported, amounting to losses of RM3.18 billion (NSRC; Mahaizura, 2024). The actual figure is likely higher due to underreporting, highlighting the urgency of effective prevention strategies. These statistics emphasize the importance of improving educational initiatives that foster informed digital behavior and protect citizens, especially vulnerable groups, from becoming victims of online scams.

This study introduces an innovative, data-driven approach to enhancing scam awareness by using machine learning techniques to identify predictors of financial loss based on victims' demographic and knowledge-related attributes. By applying decision trees and logistic regression models, this research

reveals patterns that can inform the design of targeted educational programs and digital literacy modules. While earlier studies often focused on psychological or social aspects, few have explored how advanced analytics can contribute to public education and behavior change. This research aligns with the innovation in educational research by providing practical insights that support the development of evidence-based learning tools and awareness strategies to reduce scam victimization in Malaysia.

METHODS

Research Design

This study employed both descriptive and exploratory research designs to investigate the factors contributing to financial loss among scam victims in Malaysia, with a focus on educational relevance. The descriptive approach helped identify scam trends and financial impacts, while the exploratory design tested hypotheses based on cybersecurity education and behavioral outcomes. A cross-sectional design was used to capture data at a single point in time, offering insights into scam vulnerability for future digital education efforts.

Population and Sample

The study focused on Malaysian individuals aged 18 and above who had experienced online scams. As the total population of scam victims is unknown, Cochran's formula was applied to calculate a suitable sample size. With a 5% margin of error and 95% confidence level, a minimum of 384 responses was required. A total of 394 valid responses were collected, aligning with social science research standards and providing sufficient data for educational modeling.

Sampling Method

Due to the unavailability of a formal sampling frame, non-probability sampling was used. Participants were recruited through convenience sampling on social media platforms. While this method limits generalizability due to potential bias, it enables timely data collection from actual scam victims. This practical approach offered valuable insights into victim profiles, supporting the design of targeted digital safety education programs.

Data Collection Method

Primary data were gathered using a self-administered online questionnaire via Google Forms. The survey was distributed through Facebook, WhatsApp, Telegram, and Instagram over a two-month period. Ethical approval was granted by the UiTM Seremban 3 Research Ethics Committee. Continuous sharing of the survey link helped reach a broad audience of potential respondents. This online approach aligned with the study's emphasis on addressing digital behavior through educational strategies.

Research Instrument

The questionnaire included 29 items across three sections. Section A gathered demographic details such as age, gender, education level, internet usage, urbanity, and scam type. Section B measured internet safety knowledge and reporting behavior, incorporating 12 items from Okanlawon et al. (2015) and 3 adapted from Bijwaard (2020). Section C examined the negative impacts of scams, including financial, emotional, and behavioral effects. The instrument was designed to inform the development of education programs in digital literacy and emotional resilience.

Data Analysis

Descriptive statistics were used to identify risk patterns. Predictive models, including logistic regression and decision tree algorithms, were applied to determine factors influencing financial loss. The dataset was split 70:30 for training and testing. SPSS and SAS Enterprise Miner were used for analysis. The decision tree model provided better classification accuracy and interpretability, supporting its use in informing educational campaigns and intervention strategies.

RESULTS AND DISCUSSION

Descriptive Analysis

The demographic findings highlight patterns relevant to scam awareness education. While gender differences are slight, females appear more susceptible. Older adults (55+) and those with only primary or secondary education face higher risks of financial loss, underscoring the need for targeted literacy and prevention programs. Moderate internet users are also more vulnerable, possibly due to inconsistent online behavior. Additionally, urban residents face greater scam exposure, likely from increased digital activity. These insights support the development of age-, education-, and region-specific strategies to promote digital safety and responsible online behavior.

Figure 1 further illustrates the types of scams encountered by victims. Online sales scams are the most prevalent, accounting for the highest proportion of financial loss among respondents. SMS scams also show considerable impact, followed closely by Macau scams and job-related scams. Though African scams and business email compromise cases were less common, they remain areas of concern due to the significant financial damage they can cause. These findings point to the necessity of tailoring scam education programs by scam type. For instance, educational modules that focus on safe online shopping habits and how to recognize phishing attempts via SMS can directly address the most common threats.

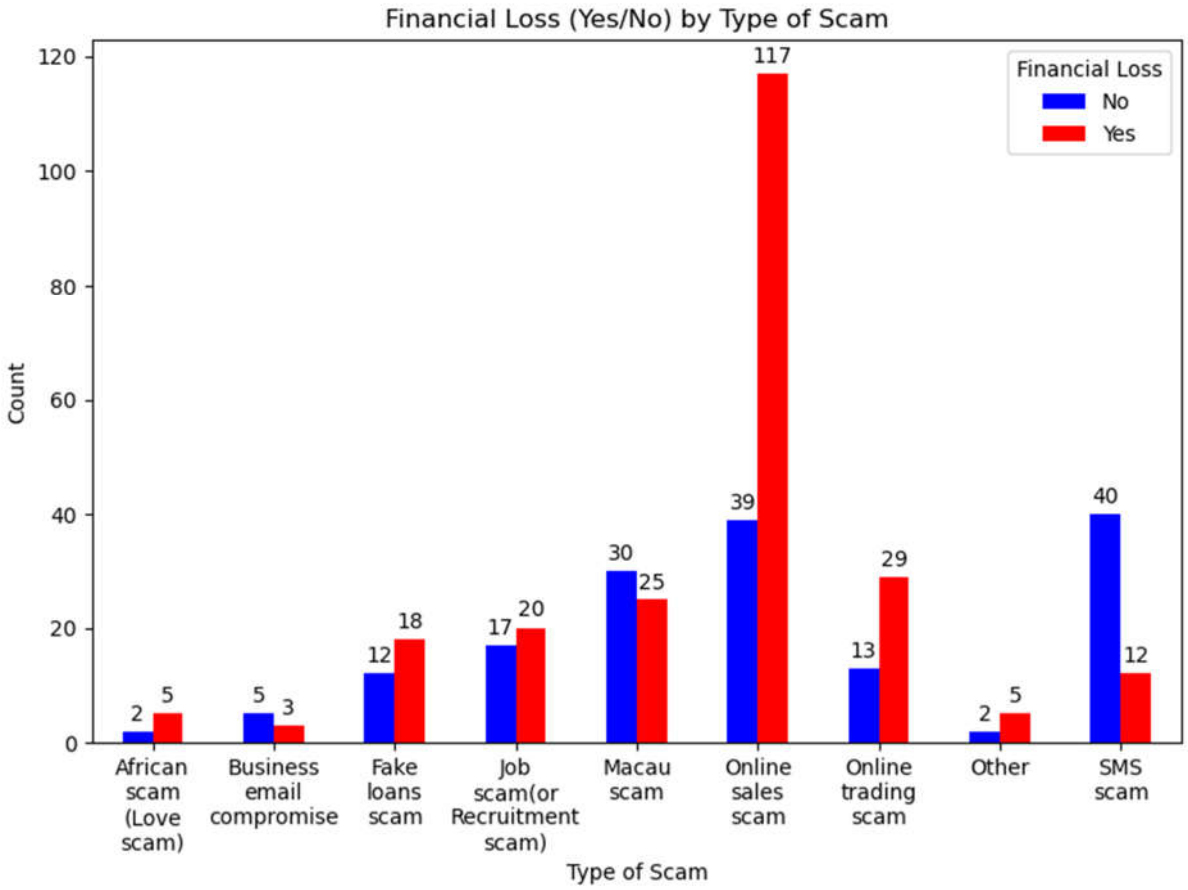

Figure 1.: Distribution of Scam Types by Financial Loss (Yes/No)

Predictive Model Performance

Table 1.: Model Performance Comparison

Comparison Metrics	**Decision Tree - GINI**	**Logistic Regression - Main Effect**
AUC	0.838	0.797
Accuracy	83.05%	72.88%
Sensitivity (Recall)	87.14%	72.86%
Specificity	72.92%	77.08%
F1 Score	85.92%	76.12%

As shown in Table 1, the decision tree model outperformed the logistic regression model across all key evaluation metrics in predicting financial loss among scam victims. The decision tree achieved an AUC value of 0.838 on the validation dataset, compared to 0.797 for logistic regression, suggesting better generalization to unseen data and stronger classification performance.

In terms of accuracy, the decision tree reached 83.05%, while logistic regression recorded 72.88%. The decision tree also showed higher sensitivity (87.14%) compared to logistic regression (72.86%), indicating superior ability to correctly identify individuals who experienced financial loss. Although logistic regression had slightly higher specificity (77.08% vs. 72.92%), the decision tree achieved a higher F1 score (85.92% vs. 76.12%), confirming its overall advantage in balancing precision and recall.

From an educational perspective, the decision tree model's interpretability and superior performance make it a practical tool for identifying high-risk individuals and guiding the development of evidence-based training programs. Its outputs can support educators, policy-makers, and digital literacy campaigners in recognizing key risk indicators and designing targeted awareness strategies accordingly.

Factors Influencing Financial Loss

Table 2.: Feature Importance

Variable	Importance Score
Emotional harm	1.0000
Knowledge score	0.5123
Age	0.3947
Gender	0.1730
Urbanity	0.1152
Level of education	0.0947
Internet usage	0.0821
Physical harm	0.0000

Table 2 shows the decision tree analysis results, highlighting the most influential factors in predicting financial loss. Emotional harm was the strongest predictor (importance = 1.0000), emphasizing the need to incorporate emotional resilience into scam prevention education. Internet safety knowledge ranked second (0.5123), reinforcing the role of digital literacy in reducing risk and the value of integrating it into learning programs.

Age followed as the third key factor (0.3947), with gender, urbanity, education level, and internet usage showing lesser influence but still relevant for tailored interventions. Physical harm had no predictive value in this context.

Overall, the results suggest that emotional vulnerability, online safety knowledge, and age are the most critical risk factors. Educational efforts should prioritize these areas to improve scam prevention, especially among high-risk groups.

CONCLUSION

This study identified emotional harm, internet safety knowledge, and age as key predictors of financial loss among scam victims in Malaysia. Among the models tested, the decision tree outperformed logistic regression, offering stronger predictive power and clearer insights into victim risk profiles.

These findings highlight the importance of data-driven approaches in educational research, particularly in designing targeted digital literacy and scam awareness programs. By focusing on emotional resilience and internet safety, educational interventions can be better tailored to vulnerable groups, especially older adults and those with lower education levels.

Overall, this research contributes to the growing field of educational innovation by demonstrating how predictive modeling can inform more effective, evidence-based learning strategies to prevent online scams.

ACKNOWLEDGEMENTS

The researcher would like to express sincere gratitude to Universiti Teknologi MARA (UiTM) Seremban 3 for the opportunity and support to conduct this study. Special thanks are extended to the UiTM Seremban 3 Research Ethics Committee for granting ethical approval and ensuring that this research was conducted responsibly.

Gratitude is also given to the respondents, especially scam victims, for their honest participation in the online survey. Appreciation is further extended to the research supervisor, lecturers, peers, and family for their guidance and support throughout this research journey.

REFERENCES

Andrade, C. (2021). The inconvenient truth about convenience and purposive samples. *Indian Journal of Psychological Medicine, 43*(1), 86–88. https://doi.org/10.1177/0253717620977000

Bhardwaj, P. (2019). Types of sampling in research. *Journal of Primary Care Specialties, 5*(3), 157–163. https://doi.org/10.4103/jpcs.jpcs_19_19

Bijwaard, D. (2020). Survey on “scams and fraud experienced by consumers”-final report.

Bornstein, M. H., Jager, J., & Putnick, D. L. (2013). Sampling in developmental science: Situations, shortcomings, solutions, and standards. *Developmental Review, 33*(4), 357–370. https://doi.org/10.1016/j.dr.2013.08.003

Brooks, R. (2022). The invention and evolution of the internet. https://online.wrexham.ac.uk/the-invention-and-evolution-of-the-internet/

Hanoch, Y., & Wood, S. (2021). The scams among us: Who falls prey and why. *Current Directions in Psychological Science, 30*(3), 260–266.

Jager, J., Putnick, D. L., & Bornstein, M. H. (2017). II. More than just convenient: The scientific merits of homogeneous convenience samples. *Monographs of the Society for Research in Child Development, 82*(2), 13–30. https://doi.org/10.1111/mono.12296

Mahaizura, A. M. (2024). Kes tipu direkod paling tinggi dalam jenayah atas talian. *Harian Metro*. https://www.hmetro.com.my/mutakhir/2024/08/1121860/kes-tipu-direkod-paling-tinggi-dalam-jenayah-atas-talian

Okanlawon, A. E., Yusuf, F. A., & Abanikannda, M. O. (2015). University students' knowledge and attitude towards internet safety: A preliminary study. *Journal of Emerging Trends in Educational Research and Policy Studies, 6*(3), 279–286.

Salkind, N. J. (2010). Encyclopedia of research design (Vol. 1). SAGE Publications.

Sedgwick, P. (2013). Convenience sampling. BMJ, 347, f6304. https://doi.org/10.1136/bmj.f6304

Sekaran, U., & Bougie, R. (2016). Research methods for business: A skill-building approach (7th ed.). John Wiley & Sons.

Whitty, M. T. (2020). Is there a scam for everyone? Psychologically profiling cyberscam, victims. *European Journal on Criminal Policy and Research, 26*(3), 399–409.