Jurnal Kejuruteraan 38(4) 2026: 1567-1575
https://doi.org/10.17576/jkukm-2026-38(4)-03

Machine Learning-Based Modelling of Health and Safety App Usability: Identifying Key Predictors in an Engineering School Setting

Fernan Patrick Flores & Yogi Tri Prasetyo*

Department of Industrial Engineering and Management,

Yuan Ze University, 135 Yuan-Tung Road, Chung-Li, 32003, Taiwan

**Corresponding author: yogi.tri.prasetyo@saturn.yzu.edu.tw*

Received 27 October 2025, Received in revised form 19 March 2026
Accepted 19 April 2026, Available online 30 July 2026

ABSTRACT

The adoption of mobile applications for occupational health and safety is growing in higher education, especially in engineering environments where students face elevated physical risks. This study explores the Perceived Usability of a health and safety mobile app among students at a Philippine engineering university. Grounded in the Unified Theory of Acceptance and Use of Technology, safety culture dimensions, and the System Usability Scale, a structured questionnaire with 48 items covering 10 variables was distributed to 224 purposively selected respondents. After data cleaning and class balancing using SMOTE, five machine learning models such as Multilayer Perceptron, Support Vector Classifier, XGBoost, LightGBM, and Random Forest Classifier were trained and tuned using grid search. RFC achieved the best performance with 90.9% accuracy and a weighted F1-score of 0.91. Validation through cross-validation, log-loss tracking, and out-of-bag error estimation confirmed the model's robustness. SHAP analysis and tree-based interpretation revealed Social Influence, Safety Training, and Safety Participation as the strongest predictors of PUS. Other contributing factors included Effort Expectancy, Performance Expectancy, Facilitating Conditions, Intention of Use, and Use Behavior. Safety Planning was excluded due to low correlation. Unlike traditional Structural Equation Modeling approaches, this machine learning framework captures complex, non-linear relationships in behavioral data and generates logic-based structures that can be embedded into rule-based systems. This study contributes to engineering systems design by providing a predictive modeling framework that supports planning and app optimization in academic environments.

Keywords: Perceived usability; health and safety; machine learning; random forest classifier; SHAP; UTAUT; SUS; mobile application; SMOTE

INTRODUCTION

Accidents and near-misses in engineering schools are not uncommon, given the daily use of heavy machinery, electrical systems, and chemical substances (Fernández-Muñiz et al., 2007). Compared to standard classrooms, these technical spaces demand constant attention and shared responsibility among students, staff, and management. Building a culture where safety becomes part of everyday decisions matters just as much as having formal rules and procedures. Yet, even with policy frameworks, unsafe behavior often persists (Griffin & Neal, 2000). Most schools rely on traditional measures like safety orientation sessions, mandated Basic Occupational Safety and Health modules, visible signage, and routine inspections (Arezes & Miguel, 2003). These tools boost awareness but don't guarantee long-term engagement. Without real-time alerts or accessible ways to report hazards, students often treat safety as a task to complete rather than a daily habit (DeJoy, 2005). This gap between policy and practice calls for more practical interventions.

Digital solutions offer promise. Industries and institutions implementing mobile apps and online hazard-reporting platforms have observed increased reporting rates, faster responses, and stronger safety engagement (Guo et al., 2017; Khosravi et al., 2014). However, most studies rely on basic surveys or interviews and rarely analyze long-term usage patterns. Moreover, research on how these tools integrate with campus life especially in developing countries is still limited. In Philippine higher education, recent investigations show that although safety systems exist, sustaining and enforcing them remains challenging (Cuya-Antonio & Gabriel, 2021). Another study links workplace safety and social support to faculty turnover in private universities, emphasizing broader implications of safety practices (Camlian & Baron, 2025). Yet, few studies examine student-centered, tech-based safety tools or apply predictive analytics to understand app usage and adoption. This signals a need for evidence-based insights into what truly influences app effectiveness.

This study seeks to fill that gap by deploying a health and safety mobile app within an engineering school and applying predictive modelling to uncover the drivers of sustained use. It's not just about testing a tool but it's about embedding safety tech into daily campus routines in a way that reflects real behavior. From a system design perspective, integrating health and safety technologies within educational institutions requires rigorous modelling to support decision-making and optimize implementation strategies. This study introduces a machine learning–based modelling framework that serves as a technical tool for predicting usability outcomes and guiding app deployment. By combining structured data inputs with advanced predictive techniques, the study offers a methodological contribution relevant to system-level planning, technological innovation, and engineering safety management.

The findings will also help administrators target efforts where they matter most and make the best use of limited resources. The main aim is to determine the factors affecting the successful implementation of a health and safety mobile app to strengthen safety culture in an engineering school setting. Specifically, this study seeks to: (1) identify and rank key drivers and barriers to app adoption; (2) apply data-driven predictive methods to reveal usage patterns; and (3) provide practical recommendations for sustainable app integration in everyday campus activities.

METHODOLOGY

RESEARCH MODEL

The proposed research model, illustrated in Figure 1, integrates elements from the Safety Culture model, the Unified Theory of Acceptance and Use of Technology (UTAUT), and usability assessment principles to explain the factors influencing the Perceived Usability (PUS) of a health and safety mobile application in an engineering school setting. Specifically, it posits that the Safety Culture dimensions, including Safety Planning (SPL), Safety Training (STR), and Safety Participation (SPR), positively influence PUS as indicated by Hypotheses 1, 2, and 3. Additionally, constructs from the UTAUT model, namely Performance Expectancy (PEX), Effort Expectancy (EEX), Facilitating Conditions (FCO), Social Influence (SIN), Intention of Use (IUS), and Use Behavior (UBE), are also hypothesized to positively and significantly affect PUS as described by Hypotheses 4 through 9. The integration of these models provides a comprehensive foundation for this study: Safety Culture offers practical safety-related dimensions grounded in Zohar's (1980) seminal work and widely applied in occupational contexts (Abella et al., 2023); UTAUT contributes well-established predictors of technology acceptance (Venkatesh et al., 2003); and the System Usability Scale (SUS), developed by Brooke (1996), ensures a standardized assessment of users' perceived system usability.

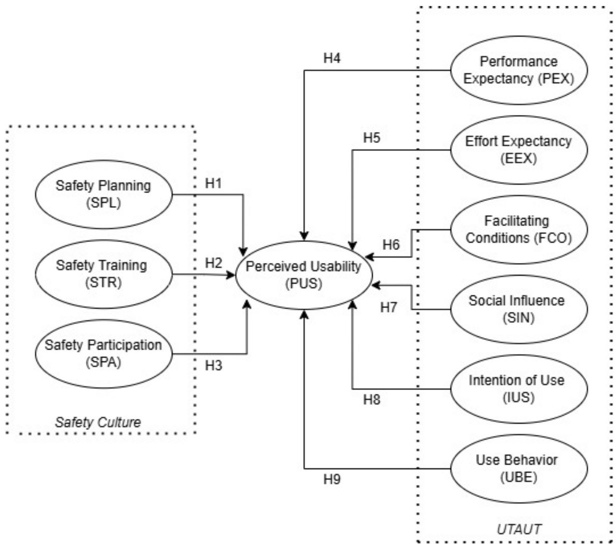

FIGURE 1. Research Model.

RESEARCH FLOW

The present study adopts a machine learning–driven modelling framework (see Figure 2) to identify key variables influencing the perceived usability of a health and safety mobile application in an engineering academic setting. It is important to note that the mobile application used in this study was pre-developed and not part of the current research's scope; hence, no system development or design process is discussed herein. This work further contributes to the technical modelling process that centers on data engineering and model optimization. The methodology began with the design and deployment of a structured questionnaire, followed by data cleaning, aggregation, label assignment, and preprocessing. To address class imbalance in the target variable, the Synthetic Minority Oversampling Technique (SMOTE) was applied (Chawla et al., 2002). Five machine learning models such as Multilayer Perceptron (MLP), Support Vector Classifier (SVC), XGBoost, LightGBM, and Random Forest Classifier (RFC) were then trained and optimized using grid search based on the highest F1-score (Shreya and Pathak, 2025; Fu et al., 2024). The best-performing configuration was evaluated through accuracy, precision, recall, and F1-score, followed by model validation using cross-validation (Stone, 1974; Rani et al., 2025), learning curves, log-loss error (Goodman et al., 1991; Rani et al., 2025), and OOB error rate for tree-based models (Breiman, 1996; Lu and Hardin, 2017). The most effective model was further analyzed using SHapley Additive exPlanations (SHAP) to extract the importance of each predictor variable (Lundberg & Lee, 2017; Fu et al., 2024). This procedure reflects similar approaches employed in prior studies that used machine learning to explore feature significance in user-centered or behavioral research contexts (Rani et al., 2025; Fu et al., 2024; Shreya and Pathak, 2025).

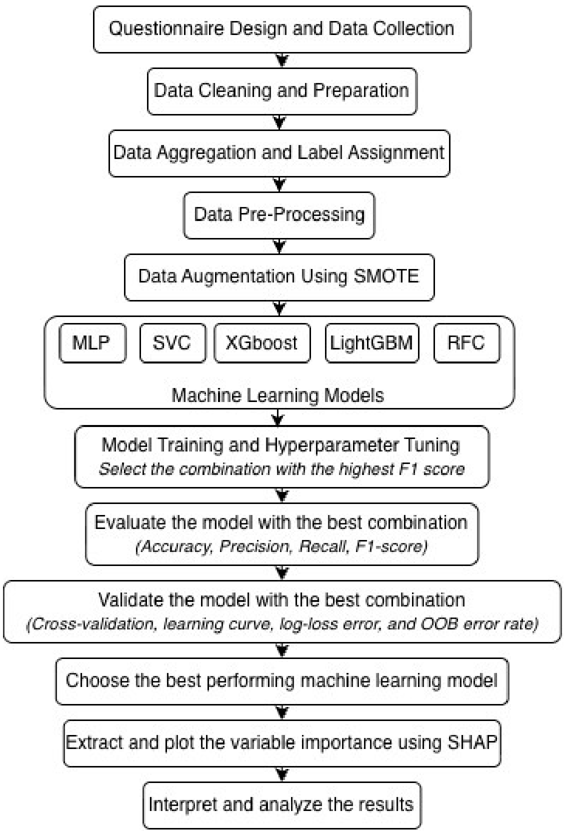

FIGURE 2. Research flow

DATA COLLECTION AND PARTICIPANTS

Data for this study were collected through a structured questionnaire comprising 48 items measuring constructs from the Unified Theory of Acceptance and Use of Technology (UTAUT), safety culture dimensions, and the System Usability Scale (SUS), using a 5-point Likert scale. Demographic questions were also included. The survey was administered online via Google Forms and followed allowed to participate. purposive sampling. Informed consent was obtained digitally, and only students who met the inclusion criteria such as current enrollment at Mapúa University, enrollment in a STEM-related undergraduate program, and prior exposure to occupational safety coursework wereBefore answering the questionnaire, participants were introduced to the application by the researcher. A total of 224 valid responses were obtained out of 232 invitations. Most respondents were male (66.51%), aged between 17 and 26 years old (98.66%, Gen Z), and primarily came from the Civil, Environmental, and Geological Engineering department (33.93%). For the full demographic distribution, refer to Table 1.

TABLE 1. Demographics (n=224)

Characteristic	Category	n	%
Status	Student	224	100.00%
Department	Civil, Environmental, and Geological Engineering (CEGE)	76	33.93%
	Mechanical and Manufacturing Engineering (MME)	55	24.55%
	Industrial Engineering and Engineering Management (IE-EMG)	39	17.41%
	Electrical, Electronics and Computer Engineering (EECE)	22	9.82%
	Chemical, Biological and Materials Engineering (CBMES)	11	4.91%
	Physics	9	4.02%
	Architecture, Industrial Design and the Built Environment (ARIDBE)	5	2.23%
Sex	Others	7	3.16%
	Male	149	66.51%
	Female	75	33.49%
Age Group	Generation Z (17–26 years old)	221	98.66%
	Millennial (31–41 years old)	3	1.34%

DATA PREPARATION AND PRE-PROCESSING

After collecting the complete responses from 224 participants, the dataset underwent a thorough cleaning and filtering process. All variables were evaluated based on their Pearson correlation with the target variable, PUS. Only indicators meeting the threshold of $r \geq 0.20$ and $p < 0.05$ were retained (Ong et al., 2022). As a result, the Safety Planning (SPL) construct was entirely excluded from the model, along with several low-correlation items across other constructs.

The remaining indicators were averaged to compute one score per construct per respondent. For the output variable, PUS (originally on a Likert scale from 1 to 5), only responses labeled 3, 4, and 5 were retained after removing underrepresented classes (Label 1 had zero instances, and Label 2 had only four). These were relabeled to 0, 1, and 2, respectively, to fit machine learning input requirements. Stratified splitting (60% train, 20% validation, 20% test) was then applied, followed by standardization using StandardScaler.

Upon inspection, the training set showed a noticeable class imbalance across the three retained labels. The original training distribution consisted of 41 samples in Class 0 (PUS = 3), 73 samples in Class 1 (PUS = 4), and only 16 samples in Class 2 (PUS = 5). This imbalance posed a risk of model bias toward the majority class. To address this issue, Synthetic Minority Oversampling Technique (SMOTE) was applied exclusively to the training set. SMOTE generated synthetic samples for the underrepresented classes (Class 0 and Class 2), resulting in an equal number of samples (73 each) across all three classes. This balanced training dataset supports fairer model learning and reduces bias toward majority class predictions, a technique widely recommended for imbalanced classification tasks (Chawla et al. 2002).

MACHINE LEARNING MODELS AND HYPERPARAMETERS

To determine the most suitable model for predicting perceived usability (PUS) based on behavioral and contextual factors, five machine learning algorithms were selected: Multi-Layer Perceptron (MLP), Support Vector Classifier (SVC), XGBoost, LightGBM, and Random Forest Classifier (RFC). These models were chosen for their diverse architectures, robustness across data types, and effectiveness in classification tasks involving user behavior and technology adoption.

The Multilayer Perceptron (MLP) Classifier, a type of feedforward artificial neural network, was selected for its ability to capture complex nonlinear relationships between input features and outcomes, especially relevant in multifactorial and behavioral datasets (Wan et al., 2018; de Oliveira et al., 2022). The Support Vector Classifier (SVC) was chosen for its effectiveness in high-dimensional feature spaces and its robust margin-based classification, which makes it suitable for both structured and imbalanced data (Zanaty, 2012). XGBoost, a widely adopted gradient boosting algorithm, was included due to its high prediction accuracy, scalability, and regularization mechanisms that help mitigate overfitting in structured datasets (Chen & Guestrin, 2016). LightGBM was utilized for its leaf-wise tree growth strategy and histogram-based optimization, which enhance computational efficiency and model performance on medium-scale datasets with complex interactions (Ke et al., 2017). Lastly, the Random Forest Classifier (RFC) was incorporated for its interpretability, robustness against overfitting, and demonstrated success in survey-based usability and safety evaluations (Chan et al., 2023).

Each model underwent grid search hyperparameter tuning using the GridSearchCV utility from the scikit-learn library. The tuning process aimed to maximize weighted F1-score, accounting for class imbalance in the multiclass classification task. Table 2 summarizes the key hyperparameters optimized per model.

TABLE 2. Selected hyperparameters for tuning.

Model	Key Parameters Tuned
MLP	hidden_layer_sizes, alpha, solver, activation
SVC	C, kernel, degree, gamma, class_weight, probability
XGBoost	n_estimators, max_depth, learning_rate, subsample, colsample_bytree, gamma, min_child_weight, reg_alpha, reg_lambda
LightGBM	n_estimators, num_leaves, learning_rate, max_depth, min_child_samples, subsample, colsample_bytree, reg_alpha, reg_lambda
RFC	n_estimators, max_depth, min_samples_split, min_samples_leaf, bootstrap, criterion

Each model underwent grid search hyperparameter tuning using the GridSearchCV utility from the scikit-learn library. The tuning process aimed to maximize weighted F1-score, accounting for class imbalance in the multiclass classification task. Table 2 summarizes the key hyperparameters optimized per model.

RESULTS AND DISCUSSIONS

HYPERPARAMETER TUNING AND EVALUATION RESULTS

All five models underwent hyperparameter tuning using grid search to identify the best-performing configuration based on weighted F1-score. The tuning process was carried out using the SMOTE-augmented training set and evaluated on the held-out test set. For the Multilayer Perceptron (MLP), the optimal configuration included a single hidden layer of 50 units, tanh activation, an alpha value of 0.001, and the lbfgs solver. The best Support Vector Classifier (SVC) used a polynomial kernel of degree 3, C=0.01, gamma=1, and no class weighting. For XGBoost, the top combination featured 200 estimators, max_depth=6, learning_rate=0.01, min_child_weight=5, and regularization terms of reg_alpha=0 and reg_lambda=2. LightGBM performed best with 200 estimators, num_leaves=15, learning_rate=0.01, and a subsample ratio of 0.8. The Random Forest Classifier achieved the highest scores with 200 trees, no limit on depth, min_samples_split=16, and bootstrap=True using the gini criterion. The final evaluation for each tuned model was based on accuracy, precision, recall, and F1-score, as summarized in Table 3.

TABLE 3. Hyperparameter tuning results.

Model	Accuracy	Precision	Recall	F1-Score
MLP	0.795	0.793	0.795	0.788
SVC	0.795	0.794	0.795	0.794
XGBoost	0.886	0.892	0.886	0.888
LightGBM	0.886	0.904	0.886	0.892
RF	0.909	0.914	0.909	0.905

The Random Forest Classifier achieved the highest overall performance with an accuracy of 90.91% and a weighted F1-score of 0.905, followed closely by XGBoost and LightGBM, both achieving 88.64% accuracy, but with slightly lower F1-scores. Meanwhile, SVC and MLP performed moderately, both with 79.55% accuracy.

VALIDATION AND SELECTION OF THE FINAL MODEL

While strong performance metrics on the test set can suggest a well-performing model, it is essential to further validate how well the model generalizes to unseen data. Specifically, assessing for overfitting and underfitting ensures that the learned patterns are not just memorized noise from the training set (Goodfellow et al. 2016). Therefore, additional validation plots and error analyses were conducted to evaluate the learning behavior of each model beyond basic evaluation scores. MLP showed perfect training performance (F1 = 1.000) but dropped on validation (F1 = 0.854), with learning curves confirming overfitting. SVC also slightly overfit, with a training F1 of 0.912 and validation F1 of 0.829. XGBoost performed well (Val F1 = 0.888) but overfit based on its widening log-loss gap. LightGBM was more stable, achieving a Val F1 of 0.892 and a smaller log-loss gap (Train = 0.354, Val = 0.528). Random Forest achieved the best balance, with Val F1 = 0.905 and stable Out-of-Bag (OOB) error (0.178), indicating good generalization. The OOB error rate plot (Figure 3) illustrates how the error stabilized as the number of trees increased, further confirming that Random Forest avoided overfitting and maintained reliable performance on unseen data.

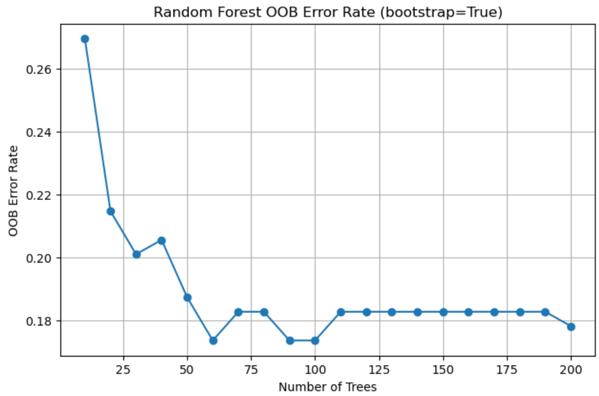

FIGURE 3. OOB Error Rate Plot.

Overall, only the Random Forest model demonstrated a strong fit without overfitting, making it the most reliable for final selection. Although XGBoost and LightGBM achieved comparable baseline accuracies, Random Forest Classifier demonstrated better generalization to unseen data. Validation tracking revealed that XGBoost was prone to overfitting, indicated by a widening log-loss gap, while LightGBM showed a smaller but still present gap. In contrast, RFC maintained a stable Out-of-Bag error rate, proving it avoided overfitting. Furthermore, RFC provides better interpretability, making it highly suitable for extracting logical decision rules in survey-based behavioral studies.

After selecting RFC as the final model, its optimum decision tree was extracted and plotted to visualize the decision logic. As shown in the optimum RFC tree plot (Figure 4), the root node of the tree is Social Influence (SIN) with a threshold of ≤ 0.51, highlighting its pivotal role as the primary splitting factor. When SIN is low (≤ 0.51), the model proceeds to assess Safety Training (STR), which further branches based on Safety Participation (SPR) and SIN itself. For instance, if STR ≤ -0.695 and SIN ≤ -0.861, the model predicts very high PUS, suggesting that strong social influence paired with basic safety training still yields confident acceptance. However, when STR and SPR are weak, the model shifts to lower or moderate predictions, particularly evident in leaf nodes with mixed class distributions.

If the initial SIN value is high (> 0.51), the tree focuses on Effort Expectancy (EEX) with a threshold of 0.124. When EEX is high (> 0.124), subsequent decisions are made based on Facilitating Conditions (FCO) and Usage Behavior (UBE). For example, if EEX > 0.81 and FCO $>$

−0.162, and UBE > 0.631, the model confidently predicts very high PUS, as seen in terminal nodes with highly skewed distributions toward that class. On the other hand, if EEX or FCO are low, the model still predicts high usability but with slightly reduced confidence, indicated by higher Gini impurity values.

Notably, variables like Performance Expectancy (PEX) and STR also appear deeper in the tree when SIN and EEX conditions are met, underscoring their nuanced but supporting influence. The decision paths collectively show that high levels of social influence, perceived effort ease, strong usage behavior, and well-supported conditions consistently lead to predictions of higher perceived usability. Conversely, when these factors are weak or inconsistent, predictions are more conservative. These branching paths offer valuable design implications for developers and administrators, providing a logic-based structure that can be embedded into rule-based systems, recommendation engines, or early-warning mechanisms within future safety app implementations.

DISCUSSION

To further interpret the internal decision logic of the Random Forest Classifier, SHapley Additive exPlanations (SHAP) was used to quantify the individual contribution of each input variable to the model's predictions (Lundberg and Lee, 2017). This technique offers both global and local explanations by calculating how each feature shifts the model's output from the base value, making it particularly useful for explaining complex ensemble models like RFC. By interpreting SHAP outputs in conjunction with the Random Forest's decision logic, this analysis contributes to a data-driven modelling framework that supports system-level decision-making. These insights enable more informed integration of usability predictors into engineering design parameters or feedback mechanisms embedded within safety applications. The final SHAP summary plot (Figure 5) illustrates the contribution of each input variable to the prediction of perceived usability (PUS).

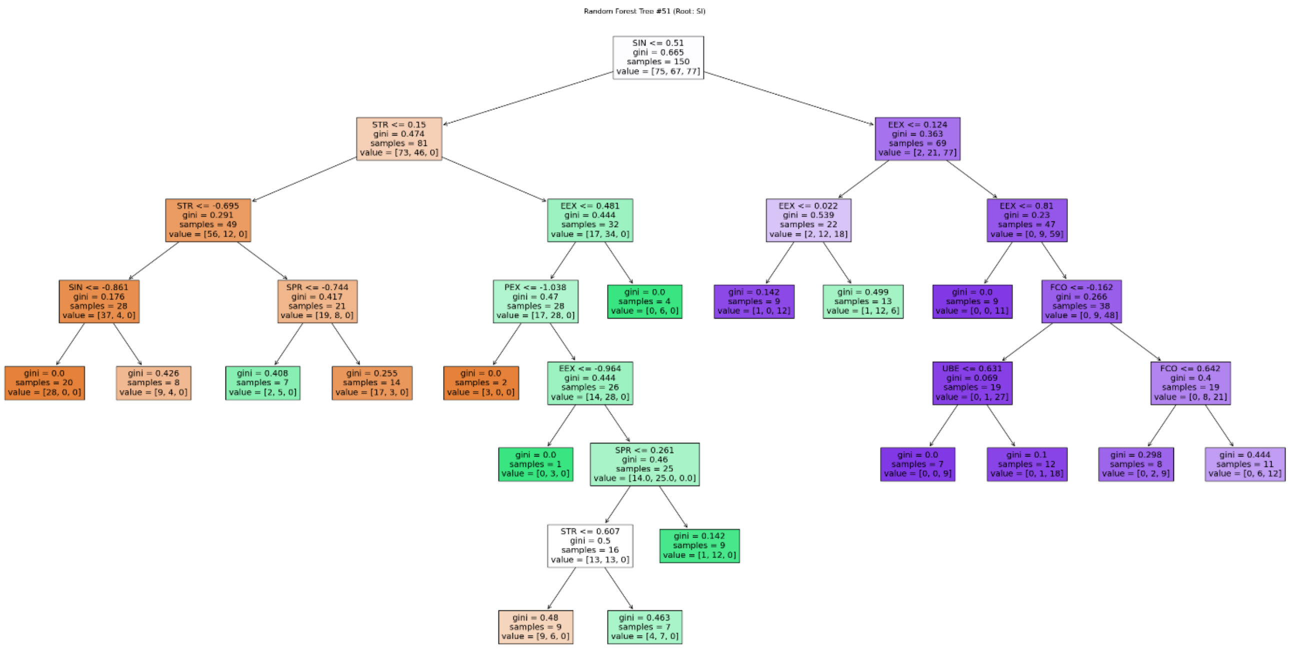

FIGURE 4. Optimum RFC Tree Plot

The analysis revealed that Social Influence (SIN) was the most significant feature in predicting Perceived Usability (PUS), with the highest importance score of 0.134. Indicators such as perceived encouragement from classmates, mentors, and influential peers toward using the health and safety app shaped how usable the system was rated. These findings echo prior work in both educational and public health settings, where social norms and endorsement positively influenced users' perceptions of usefulness and ease of use (Shen et al. 2006; Haverila et al. 2024).

Safety Training (STR) followed with an importance score of 0.069, indicating that adequate instruction and training provided by the university played a critical role in shaping usability perceptions. When students and staff felt confident in their ability to assess hazards and navigate safety procedures due to training, they were more likely to view the app as user-friendly. These results are supported by Wong et al. (2021), who found that safety training enhanced users' perceived ease of use and usefulness in safety-related technologies. Safety Participation (SPR) had a normalized importance of 0.054, suggesting that users who actively engage in voluntary safety behaviors, such as promoting the app or reporting hazards, found the system more usable. As seen in workplace safety research by Hu et al. (2016), discretionary safety behaviors are often reinforced by systems perceived as valuable and easy to use, reinforcing the idea that engagement fosters usability.

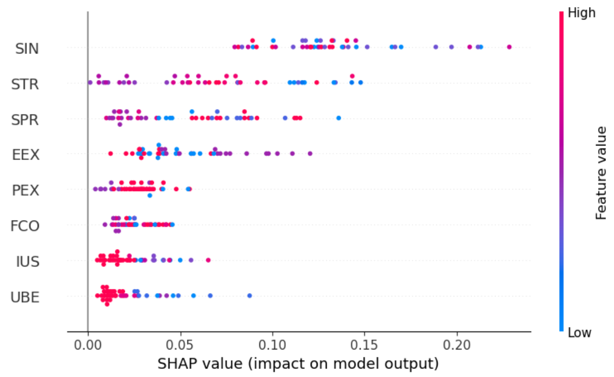

FIGURE 5. Variable importance analysis

Effort Expectancy (EEX) contributed to usability with an importance score of 0.052, highlighting the role of perceived ease of use in shaping user experience. Students who found the app simple to learn and operate tended to view it as more usable overall. This finding aligns with Ong et al. (2022), who emphasized the importance of minimizing user effort in mobile app design, especially in safety and utility contexts. Performance Expectancy (PEX) scored 0.027 in importance, indicating that students who believed the app would improve their ability to accomplish safety-related tasks perceived it as more usable. Features such as reporting hazards beyond working hours or regular digital logging of reports supported this expectation, as also reflected in findings by Ong et al. (2022).

Facilitating Conditions (FCO) held an importance score of 0.026, showing that access to resources, skills, and compatible technologies was moderately influential in shaping usability perceptions. Students who felt technically supported or believed the app was compatible with their devices reported higher PUS. This is consistent with findings from Wong (2015) and Ebadi and Raygan (2023), which emphasize the value of environmental support in technology acceptance.

Intention to Use (IUS) yielded an importance score of 0.023, indicating that stronger behavioral intention toward using the app was linked to higher PUS. This may be due to greater exposure and mental preparedness, making the system feel more intuitive. Prasetyo et al. (2025) observed a similar relationship in the context of mobile payment systems, where intent to use significantly boosted perceived usability. Use Behavior (UBE), with an importance score of 0.022, still showed a positive contribution to PUS. Students who reported actual usage of the app whether to file reports, access safety information, or improve awareness were more likely to view it as usable. This supports Lee and Koubek (2010), who found that actual use reinforces familiarity and ease, thereby improving perceived system usability.

Lastly, Safety Planning (SPL) was excluded from the final model due to weak correlation with PUS, consistent with limited evidence linking planning activities directly to usability perceptions. Previous studies such as Dvir et al. (2003) also noted that implementation of planning processes does not necessarily translate to higher satisfaction or successful technology engagement, which may explain its low predictive value.

The results provide useful insights for improving the implementation of safety-related mobile applications in academic or workplace settings. Furthermore, these findings offer direct implications for engineering system design and safety management. Rather than treating these behavioral factors as isolated metrics, administrators can integrate them directly into safety management protocols. For example, the strong predictive power of Social Influence and Safety Training dictates that successful app deployment requires structured social onboarding and direct integration into engineering curriculums, rather than passive availability. From a technical standpoint, these findings can guide the iterative development of safety app modules and inform adaptive system features such as in-app nudges, usage-based customization, or automated alert routing. This helps transition the app from a static reporting tool into a dynamic, intelligent safety assistant.

Future studies should consider expanding the respondent pool to improve model generalizability and support more robust training of machine learning algorithms. Some models in this study showed limitations likely due to the relatively small sample size, which can affect prediction stability and variable interpretation. Additionally, once the health and safety app is fully deployed and accessible across different campuses, researchers are encouraged to explore usability perceptions in various academic institutions in the Philippines. This could reveal context-specific factors or common usability trends that would strengthen the framework's applicability. Researchers may also consider longitudinal studies to assess how perceived usability evolves with continued exposure and actual use of the app over time. Integrating qualitative feedback, such as interviews or focus groups, can provide deeper insights into user experience and system improvements.

CONCLUSION

This study investigated the Perceived Usability (PUS) of a health and safety mobile application in a Philippine engineering university using a machine learning approach. Five models were compared, with the Random Forest Classifier (RFC) showing the highest accuracy and generalization performance based on cross-validation, out-of-bag error, and SHAP analysis. Among the predictors, Social Influence (SIN), Safety Training (STR), and Safety Participation (SPR) emerged as the most influential variables. Other constructs such as Effort Expectancy (EEX), Performance Expectancy (PEX), Facilitating Conditions (FCO), Intention of Use (IUS), and Use Behavior (UBE) also contributed positively to PUS, although with lower impact. Safety Planning (SPL) was excluded from the final model due to weak correlation and limited predictive value. These results emphasize the role of social and behavioral dynamics in shaping usability perceptions of safety apps, while also contributing to a predictive modelling framework applicable to human-centered system design. The combined use of advanced classification algorithms and SHAP-based interpretation provides an evidence-driven foundation for optimizing app features, user experience, and adaptive safety workflows. This work supports engineering management efforts to embed smart, data-informed tools into safety-critical environments. Furthermore, this modelling framework can be directly transferred and adapted by other engineering institutions to evaluate and deploy their own safety technologies which will allow them to identify the key behavioral and design factors essential for efficient app usability.

ACKNOWLEDGEMENT

Authors would like to thank Yuan Ze University for supporting this work.

DECLARATION OF COMPETING INTEREST

None.

REFERENCES

Abella, A. A., Prasetyo, Y. T., Young, M. N., Nadlifatin, R., Persada, S. F., Redi, A. A. N. P. & Chuenyindee, T. 2023. The effect of positive reinforcement of behavioral-based safety on safety participation in Philippine coal-fired power plant workers: A partial least squares structural equation modeling approach. *International Journal of Occupational Safety and Ergonomics* 29(3): 951–962.

Arezes, P. M. & Miguel, A. S. 2003. The role of safety culture in safety performance measurement. *Measuring Business Excellence* 7(4): 20–28.

Breiman, L. 1996. Out-of-bag estimation. Technical report, Department of Statistics, University of California, Berkeley. Retrieved from https://www.stat.berkeley.edu/~breiman/OOBestimation.pdf

Brooke, J. 1996. SUS: A 'quick and dirty' usability scale. In Jordan, P. W., Thomas, B., Weerdmeester, B. A. & McClelland, A. L. (eds.). *Usability Evaluation in Industry*. London: Taylor & Francis.

Camlian, M. M. & Baron, J. V. 2025. Workplace health and safety, social support, and turnover intention in private higher education institutions in the Philippines. *Annals of Human Resource Management Research* 5(1): 1–14.

Chan, R. Y.-Y., Wong, C. M. V. & Yum, Y. N. 2023. Predicting behavior change in students with special education needs using multimodal learning analytics. *IEEE Access* 11: 63238–63251.

Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. 2002. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16(1): 321–357.

Chen, T. & Guestrin, C. 2016. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. New York: Association for Computing Machinery.

Cuya-Antonio, O. C. M. & Gabriel, A. G. 2021. Analysis of the occupational safety and health practices of public organization in the Philippines. *International Journal of Management* 12(1): 175–190.

DeJoy, D. M. 2005. Behavior change versus culture change: Divergent approaches to managing workplace safety. *Safety Science* 43(2): 105–129.

de Oliveira, W. G., Filho, P. P. R. & da Silva Junior, E. T. 2022. Driver behavior analysis: Abnormal driving detection using MLP classifier applied to outdoor camera images. In *Intelligent Systems Design and Applications*. *Lecture Notes in Networks and Systems*, edited by Abraham, A. et al. 418. Cham: Springer.

Dvir, D., Raz, T. & Shenhar, A. J. 2003. An empirical analysis of the relationship between project planning and project success. *International Journal of Project Management* 21(2): 89–95.

Ebadi, S. & Raygan, A. 2023. Investigating the facilitating conditions, perceived ease of use and usefulness of mobile-assisted language learning. *Smart Learning Environments* 10: 30.

Fernández-Muñiz, B., Montes-Peón, J. M. & Vázquez-Ordás, C. J. 2007. Safety management system: Development and validation of a multidimensional scale. *Journal of Loss Prevention in the Process Industries* 20(1): 52–68.

Fu, Q., Wu, Y., Zhu, M., Xia, Y., Yu, Q., Liu, Z., Ma, X. & Yang, R. 2024. Identifying cardiovascular disease risk in the U.S. population using environmental volatile organic compounds exposure: A machine learning predictive model based on the SHAP methodology. *Ecotoxicology and Environmental Safety* 286: 117210.

Goodfellow, I., Bengio, Y. & Courville, A. 2016. *Deep Learning*. Cambridge, MA: MIT Press.

Griffin, M. A. & Neal, A. 2000. Perceptions of safety at work: A framework for linking safety climate to safety performance, knowledge, and motivation. *Journal of Occupational Health Psychology* 5(3): 347–358.

Guo, B. H. W., Yiu, T. W. & González, V. A. 2016. Predicting safety behavior in the construction industry: Development and test of an integrative model. *Safety Science* 84: 1–11.

Haverila, M. J., Haverila, K. & McLaughlin, C. 2024. The impact of perceived effectiveness of non-pharmaceutical interventions (NPIs) on attitude toward usage, behavioral intentions, and actual usage. *SAGE Open* 14(2): 21582440241253360.

Hu, X., Griffin, M. A. & Bertuleit, M. 2016. Modelling antecedents of safety compliance: Incorporating theory from the technological acceptance model. *Safety Science* 87: 292–298.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W. & Liu, T. Y. 2017. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems* 30: 3146–3154. Red Hook, NY: Curran Associates, Inc.

Khosravi, Y., Asilian-Mahabadi, H., Hajizadeh, E., Hassanzadeh-Rangi, N., Bastani, H. & Behzadan, A. H. 2014. Factors influencing unsafe behaviors and accidents on construction sites: A review. *International Journal of Occupational Safety and Ergonomics* 20(1): 111–125.

Lee, S. & Koubek, R. J. 2010. Understanding user preferences based on usability and aesthetics before and after actual use. *Interacting with Computers* 22(6): 530–543.

Lu, B. & Hardin, J. 2017. Constructing prediction intervals for random forests. Doctoral dissertation, Pomona College.

Lundberg, S. M. & Lee, S.-I. 2017. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* 30: 4765–4774. Red Hook, NY: Curran Associates, Inc.

Ong, A. K. S., Prasetyo, Y. T., Kusonwattana, P., Mariñas, K. A., Yuduang, N., Chuenyindee, T., Robas, K. P. E., Persada, S. F. & Nadlifatin, R. 2022. Determining factors affecting the perceived usability of air pollution detection mobile application 'AirVisual' in Thailand: A structural equation model forest classifier approach. *Heliyon* 8(12): e12538.

Prasetyo, Y. T., Susanto, K. C., Chuang, K. H., Yin, R. T., Chen, J. W., Zhang, Y. X., Benito, O. P., Belmonte, Z. J. A., Cahigas, M. M. L., Nadlifatin, R. & Gumasing, M. J. J. 2025. Factors influencing the perceived usability of Line Pay: An extended technology acceptance model approach. *Acta Psychologica* 255: 104924.

Rani, R., Jaiswal, G., Nancy, Lipika, Bhushan, S., Ullah, F., Singh, P. & Diwakar, M. 2025. Enhancing liver disease diagnosis with hybrid SMOTE-ENN balanced machine learning models: An empirical analysis of Indian patient liver disease datasets. *Frontiers in Medicine* 12: 1502749.

Shen, D., Laffey, J., Lin, Y. & Huang, X. 2006. Social influence for perceived usefulness and ease-of-use of course delivery systems. *Journal of Interactive Online Learning* 5(3): 270–282.

Shreya & Pathak, H. 2025. Explainable artificial intelligence credit risk assessment using machine learning. *arXiv* preprint arXiv:2506.19383.

Stone, M. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)* 36(2): 111–133.

Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D. 2003. User acceptance of information technology: Toward a unified view. *MIS Quarterly* 27(3): 425–478.

Wan, S., Liang, Y., Zhang, Y. & Guizani, M. 2018. Deep multi-layer perceptron classifier for behavior analysis to estimate Parkinson's disease severity using smartphones. *IEEE Access* 6: 36825–36833.

Wong, G. K. W. 2015. Understanding technology acceptance in pre-service teachers of primary mathematics in Hong Kong. *Australasian Journal of Educational Technology* 31(6).

Wong, T. K. M., Man, S. S. & Chan, A. H. S. 2021. Exploring the acceptance of PPE by construction workers: An extension of the technology acceptance model with safety management practices and safety consciousness. *Safety Science* 139: 105239.

Zanaty, E. A. 2012. Support vector machines (SVMs) versus multilayer perceptron (MLP) in data classification. *Egyptian Informatics Journal* 13(3): 177–183.

Zohar, D. 1980. Safety climate in industrial organizations: Theoretical and applied implications. *Journal of Applied Psychology* 65(1): 96–102.