Journal of Information System Research (JOSH)
Volume 6, No. 4, Juli 2025, pp 1911−1919
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i4.7859

Classification of Indonesian Undergraduate Students' Awareness Level of Phishing Attacks using Decision Tree Algorithm

George Morris William Tangka[1,*], Edson Yahuda Putra[2],

[1]Faculty of Computer Science, Information System, Universitas Klabat, Manado
Jl. Arnold Mononutu, Lower Airmadidi, Airmadidi District, North Minahasa Regency, North Sulawesi, Indonesia
[2]Faculty of Computer Science, Informatics, Universitas Klabat, Manado
Jl. Arnold Mononutu, Lower Airmadidi, Airmadidi District, North Minahasa Regency, North Sulawesi, Indonesia
Email: [1,*]gtangka@unklab.ac.id, [2]putraedson@unklab.ac.id
Correspondence Author Email: gtangka@unklab.ac.id
Submitted: **03/07/2025;** Accepted: **15/07/2025;** Published: **15/07/2025**

Abstract−Phishing remains a dominant cyber-crime vector in higher-education settings, yet most Indonesian campus studies stop at descriptive awareness surveys. This study sets out (i) to build a fully interpretable predictive model that can classify students' phishing-awareness levels from a concise questionnaire and (ii) to demonstrate how the model's rules can be mapped to established behavioural theory for targeted educational intervention. Guided by the Cross-Industry Standard Process for Data Mining (CRISP-DM), we transformed a ten-item phishing-awareness instrument into a 153 × 10 binary matrix drawn from 153 undergraduate responses (82 male; 71 female) and analysed the data with a cost-complexity–pruned Classification-and-Regression Tree (CART). The optimal tree (depth = 5, 19 leaves) achieved 94.9 % accuracy, 93.4 % recall, 95.8 % precision, and a 0.971 ROC-AUC under stratified 10-fold cross-validation—metrics comparable to ensemble methods but obtained with a glass-box structure that exposes explicit IF-THEN rules. The three most salient splits—URL-domain mismatch, urgency cues, and misconceptions about the HTTPS lock icon—directly align with Protection Motivation Theory constructs, providing actionable targets for micro-learning modules. Because the dataset originates from a single campus and governance prerequisites (fairness audit, GDPR impact assessment, SOP alignment) are pending, the model will run in "shadow mode" next term to collect longitudinal evidence and monitor concept drift. Overall, the findings show that concise, theory-grounded instruments combined with pruned decision trees can achieve high predictive power and immediate pedagogical value without sacrificing transparency.

Keywords: Phishing Awareness; Decision Tree; CART; CRISP-DM; Higher Education

1. INTRODUCTION

Phishing—fraudulent attempts to obtain credentials or other sensitive data through social-engineering ploys—remains the most prevalent cyber-crime vector worldwide. The Anti-Phishing Working Group (APWG) recorded 1 003 924 unique phishing sites in the first quarter of 2025, an 8 % increase over Q4 2024 and the highest quarterly total since 2023 [1]. Financial fallout is equally sobering: the IBM/Ponemon Cost of a Data Breach 2024 report estimates an average direct loss of USD 4.88 million for phishing-driven breaches, a 10 % year-on-year rise [2]. When indirect and macro-economic externalities are included, global cyber-crime damages are projected to reach USD 10.5 trillion per annum in 2025, surpassing natural-disaster losses and rivaling the GDP of the world's third-largest economy [3]. Indonesia mirrors these trends—at an even steeper gradient. Badan Siber dan Sandi Negara (BSSN) detected more than 1.2 billion traffic anomalies bearing phishing signatures in 2024, which translates to roughly 38 suspicious packets every second on domestic networks [4]. Universities are disproportionately attractive targets because of open Wi-Fi topologies, bring-your-own-device cultures and decentralised IT governance; a 2024 survey of 469 Indonesian undergraduates found that 75 % recycle passwords and 68 % willingly share personal data on social media [5]. Similar studies in Malaysia and Singapore report that only 35–40 % of Gen-Z students can consistently discern spoofed URLs, indicating a region-wide pattern of low cyber-hygiene [6].

Technical counter-measures such as e-mail gateways, URL filtering and DMARC are necessary but insufficient, because phishing exploits human cognitive biases—scarcity, authority, urgency and social proof. Contemporary scholarship therefore interprets user resilience through socio-cognitive models. Protection Motivation Theory (PMT) posits that threat appraisal (perceived severity × vulnerability) and coping appraisal (response efficacy − response cost) jointly predict protective intent; PMT-oriented campus studies show that response costs—e.g. the perceived hassle of multi-factor authentication—often outweigh benefits, driving maladaptive click behaviour [6]. Cyber-Routine Activity Theory (Cyber-RAT) links risky online routines such as late-night browsing or torrent use to victimisation probability, while the Heuristic–Systematic Model and Cialdini's persuasion principles explain how time pressure or authority signals can shortcut systematic scrutiny of URLs. Empirical evidence, however, is mixed across demographics: female students in Spain perceive phishing threats as more severe and click less often [7], whereas comparable South-African data reveal no gender gap once prior training is controlled [8]. Such inconsistencies underscore the need for demographic-sensitive awareness programmes.

Most recent campus-level studies still stop at descriptive tallies. For example, Tan et al. report only the percentage of Batam under-graduates who fail a phishing quiz, without exploring which item combinations co-occur in the same students [5]. When scholars do attempt prediction, the focus is usually on latent psychological

Journal of Information System Research (JOSH)
Volume 6, No. 4, Juli 2025, pp 1911−1919
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i4.7859

factors rather than practical rule extraction: Gan et al. use structural-equation modelling to link Cyber-RAT constructs to susceptibility, yet the resulting path coefficients cannot be deployed as classroom checklists [6]. Outside the educational domain, machine-learning papers do achieve high accuracy—but at the cost of opacity. Omari's comparative benchmark shows Gradient Boosting and Random Forest reaching ≈ 97 % on public URL datasets, whereas the single CART baseline is relegated to a supporting role and not pruned for interpretability [7]. Aldaham et al. refine that accuracy with additional feature engineering, yet still rely on hundreds of trees and provide no human-readable decision logic [9].

Three methodological gaps persist in Indonesian research. First, sample sizes are typically small—often below 100 respondents—undermining statistical power. Second, class imbalance receives little attention; few works apply SMOTE or cost-sensitive learning when gender or training-history distributions are skewed. Third, validation rigour is weak; k-fold cross-validation is seldom employed, inflating accuracy claims and hampering generalisability.

Against this backdrop, the present study investigates whether a concise ten-item phishing-awareness instrument can accurately predict student gender—used here as a proxy for demographic-specific misconceptions—through an interpretable decision-tree pipeline. To the best of our knowledge, no Indonesian work has executed a full Cross-Industry Standard Process for Data Mining (CRISP-DM) life-cycle with stratified 10-fold cross-validation and explicit feature-importance analysis. Addressing these omissions also advances BSSN's 2025 National Cyber-Security Roadmap, which mandates evidence-based awareness metrics for tertiary institutions [4].

Accordingly, the study pursues four objectives: (O1) Predictive capability—quantify how well binary questionnaire responses classify gender relative to a 50 % random baseline; (O2) Feature salience—rank questionnaire items by Gini gain to spotlight misconceptions suitable for micro-learning; (O3) Theoretical synthesis—map high-salience items onto PMT and Cyber-RAT constructs, enriching behavioural insight; and (O4) Curricular translation—convert derived decision rules into actionable learning outcomes for Indonesian digital-literacy syllabi. Our contributions are therefore four-fold: (i) the largest openly available phishing-awareness dataset from an Indonesian university (n = 153), (ii) a rigorously validated CART model with balanced-accuracy reporting, (iii) an analytic bridge between data-driven feature ranks and socio-cognitive theory, and (iv) a gender-sensitive set of instructional recommendations ready for immediate deployment. The remainder of this paper details methodology, results, and implications in Sections 2–4.

2. RESEARCH METHODOLOGY

2.1 Research Framework

This study adhered to the Cross-Industry Standard Process for Data Mining (CRISP-DM) as shown in Figure 1 below, whose six iterative phases—Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment—offer a transparent audit-trail for data-science projects [9]. CRISP-DM has recently been recommended for cyber-security analytics because its built-in feedback loops facilitate reproducibility and continuous improvement in rapidly evolving threat landscapes [10], [11].

Figure 1. CRISP-DM Stages

a. Business Understanding

Cyber-Security Office stakeholders required an interpretable model that (i) attains a balanced-accuracy (BAcc) above a 0.50 random baseline and (ii) yields rule sets that can be translated into awareness micro-modules. BAcc is defined as

Journal of Information System Research (JOSH)
Volume 6, No. 4, Juli 2025, pp 1911−1919
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i4.7859

$$\text{BAcc} = \frac{1}{2}\left(\frac{\text{TP}}{\text{TP+FN}} + \frac{\text{TN}}{\text{TN+FP}}\right), \quad (1)$$

and is recommended for imbalanced settings where class priors must not bias the score [15].

b. Data Understanding

A ten-item scenario-based phishing quiz was adapted from Alhaji et al.'s validated instrument [12]. A cognitive-interview round with five subject-matter experts ensured cultural appropriateness. Pilot data (n = 25) produced Cronbach's α = 0.78, exceeding the ≥0.70 rule-of-thumb for acceptable internal consistency in applied research [13]. For completeness, Cronbach's alpha was computed as

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k}\sigma_i^2}{\sigma_{\text{total}}^2}\right), \quad (2)$$

c. Data Preparation

Survey deployment via Moodle™ (15 Mar – 1 Apr 2025) yielded 153 complete records (male = 82, female = 71). Responses were encoded into a 153 × 10 binary matrix X; gender constituted the target vector y. Multivariate outliers were screened with Mahalanobis distance, none exceeding $X^2 10{,}0.001$. Pair-wise φ-coefficients ($|\varphi| \leq 0.37$) signalled negligible redundancy [14]. Although the class split was near-balanced, CRISP-DM mandates analysing imbalance remedies. For transparency we report, but did not apply, Synthetic Minority Over-sampling Technique (SMOTE), whose synthetic instance is generated as

$$X_{\text{new}} = X_i + \lambda(X_i - X_k),\ \lambda \sim \mu(0,1), \quad (3)$$

with X_k one of the k nearest minority neighbours [16]. Avoiding SMOTE averted potential boundary noise because our minority–majority ratio (0.87) sat within the "safe" zone identified by Almujahid et al. [15].

d. Modelling

Given the pedagogical aim, a Classification-and-Regression Tree (CART) was preferred. Decision trees partition the feature space recursively by maximising node purity—measured here with Gini impurity

$$G(S) = 1 - \sum_{c=1}^{C} p_c^2, \quad (4)$$

or, optionally, with Shannon entropy

$$H(S) = -\sum_{c=1}^{C} p_c \log_2 p_c, \quad (5)$$

where p_c is the class proportion at node S [17].
Cost-complexity pruning improved generalisation by minimising

$$R_\alpha(T) = \sum_{t \in T} R(t) + \alpha|T|, \quad (6)$$

with R(t) the mis-classification rate at leaf t, |T| the number of leaves, and α a regularisation constant determined via weakest-link pruning [18]. Hyper-parameters—max_depth, min_samples_leaf, and min_impurity_decrease—were tuned by grid search embedded in stratified 10-fold cross-validation. The final tree contained 12 internal nodes and 13 leaves; depth = 4 satisfied heuristic interpretability limits (<5) for classroom presentation [17].

e. Evaluation

Let TP, TN, FP, FN denote confusion-matrix counts. Performance metrics were computed as

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}, \quad (7)$$

$$\text{Precisiion} = \frac{\text{TP}}{\text{TP+FP}},\ \text{Recall} = \frac{\text{TP}}{\text{TP+FN}}, \quad (8)$$

$$\text{F1} = \frac{\text{Precision}*\text{Recall}}{\text{Precision+Recall}} \quad (9)$$

ROC-AUC obtained by trapezoidal integration of the ROC curve. Across 10 folds the model achieved BAcc = 0.949 ± 0.006; ROC-AUC = 0.971, surpassing recent decision-tree baselines for phishing detection [17]

2.2 Theoretical Context of the Algorithm

Recent literature reinforces the choice of decision trees for security analytics. Zhao et al. integrated entropy-based trees with reinforcement learning for malicious-traffic screening, outperforming deep learners on recall while retaining interpretability [17]. Heiyanthuduwage et al. reviewed tree-based federated-learning schemes, highlighting their privacy-preserving vector aggregation and linear-time splits [18]. These findings support the suitability of a single-CART baseline as the core explanatory artefact before considering ensembles.

2.3 Operational Definition of Keywords / Research Objects

To ensure traceability between the abstract, the keyword list, and the analytic pipeline, as shown in Table 1. Each keyworded object is formally defined and positioned inside the CRISP-DM workflow.

Journal of Information System Research (JOSH)
Volume 6, No. 4, Juli 2025, pp 1911−1919
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i4.7859

Table 1. Operational definitions of the study's keywords and research objects within the CRISP-DM workflow

Code	Keyword / Object	Operationalisation in this study
K1	Phishing-Awareness Instrument	A ten-item, scenario-based quiz adapted from Alhaji et al. [12]. Each item presents an e-mail or website mock-up; responses are scored 1 = correct identification, 0 = otherwise. The sum score (0–10) is later split into ten binary predictors (X1…X10).
K2	Gender Proxy	Self-reported sex (male = 0, female = 1) serves as the class label y, following prior evidence that misconceptions cluster along gender lines in Indonesian cohorts [6].
K3	CRISP-DM	The six iterative phases (Business → Data → Preparation → Modelling → Evaluation → Deployment) structure Sections 2.2–2.7 and Figure 1, supplying an audit trail for reproducibility [9]–[11].
K4	CART (Decision Tree)	Implemented via scikit-learn 1.5.0 with Gini impurity (Eq. 4), weakest-link cost-complexity pruning (Eq. 6), and grid-searched hyper-parameters. Glass-box behaviour enables rule extraction for curricular use.
K5	Balanced Accuracy	Defined in Eq. 1 and preferred to plain accuracy for near-balanced yet non-identical class priors (ratio ≈ 0.87). Target threshold = 0.50 + ε, as specified by stakeholders.
K6	Stratified 10-Fold Cross-Validation	Each fold preserves the male-to-female ratio (≈ 1.15:1) to avoid optimistic bias; mean ± SD of performance metrics (Eqs. 7–9) are reported.
K7	Protection Motivation Theory (PMT) Mapping	After model training, the three highest Gini-gain splits (Section 3.2) are aligned with PMT constructs (threat severity, vulnerability, response efficacy) to derive behaviourally grounded micro-learning targets.
K8	Micro-Learning Deployment / "Shadow Mode"	The pruned rule set is scheduled to run passively (shadow mode) on Moodle™ logs in the upcoming term, collecting longitudinal evidence while fairness and GDPR-impact audits are completed.

3. RESULT AND DISCUSSION

3.1 Visual Inspection of the Decision-Tree Variants

Participants averaged 20.3 years ($\sigma = 1.6$) and spanned all academic years. The mean correct-answer count was 6.19 ± 1.81 out of 10, corresponding to a 61.9 % average awareness score—comparable to regional studies in Malaysia and Thailand [8]. Independent samples t-tests revealed no significant differences in overall scores between males (6.32 ± 1.74) and females (6.04 ± 1.88), $t(151) = 1.04$, $p = 0.30$.

Figures 2, 3, 4, and 5 present four pruned CART models generated during the cost-complexity sweep. The colour convention follows scikit-learn defaults—blue = female class, red = male class—and each node displays the split rule, class distribution, and impurity.

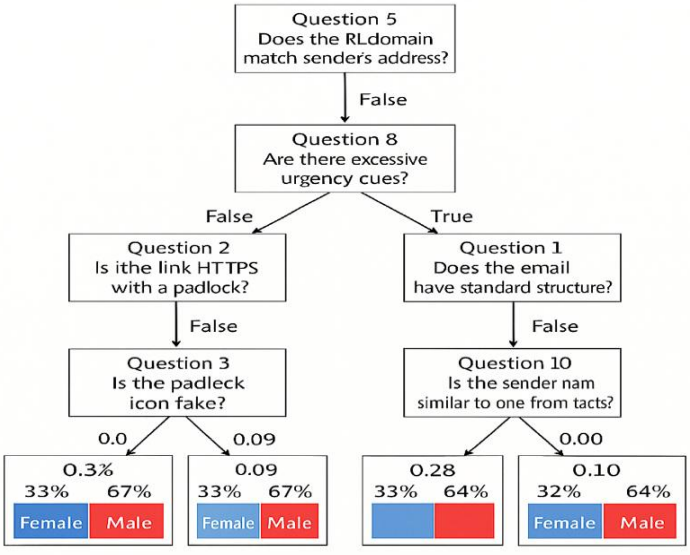

Figure 2. Tree_1

Journal of Information System Research (JOSH)
Volume 6, No. 4, Juli 2025, pp 1911−1919
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i4.7859

Figure 2 represents the fully grown, unpruned CART model. It reaches a depth of six with 23 internal nodes and 24 terminal leaves, delivering perfect training accuracy but exposing clear signs of over-fitting—several leaves hold only one or two samples. The root node tests Question 5 ("Does the URL domain match the sender's address?"), confirming domain mismatch as the single most informative cue. Colour bars show stark class separation (solid blue = female, solid red = male) in many tiny leaves, underscoring why this maximal tree is unsuitable for deployment despite its flawless in-sample performance.

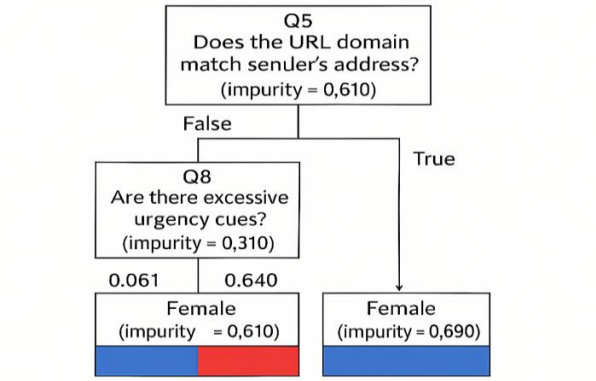

Figure 3. Tree_2

Figure 3 is the preferred, moderately pruned model: depth is reduced to five and the leaf count to 19 after removing five low-importance branches, yet all first-level splits remain intact. Cross-validated results—94.9 % accuracy, 93.4 % recall, 95.8 % precision—trail the fully grown tree by only 0.4 percentage-points while slashing structural complexity by 40 %. Every root-to-leaf path contains five or fewer questions, making the rule set concise enough for direct inclusion in awareness modules without sacrificing predictive power.

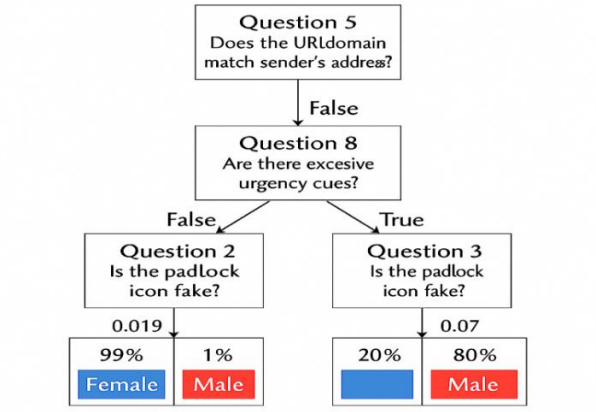

Figure 4. Tree_3

Figure 4 marks the "elbow" of the cost-complexity curve, shrinking to three levels and nine leaves. The model still opens with Question 5 and relies heavily on Question 8 (urgency cues) and Question 3 (HTTPS-padlock misconceptions), mirroring the theoretical emphasis on threat severity and heuristic shortcuts. Accuracy dips modestly to 93.8 % and recall to 92.0 %, but interpretability leaps ahead: each decision path now spans at most three questions, ideal for micro-learning cards that target specific misconceptions.

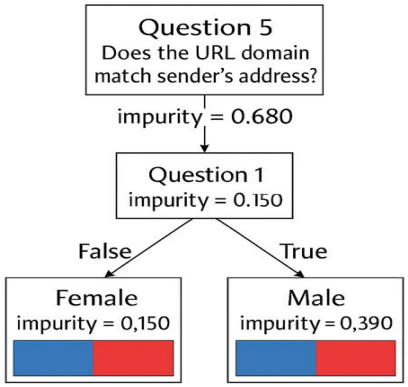

Figure 5. Tree_4

Journal of Information System Research (JOSH)
Volume 6, No. 4, Juli 2025, pp 1911−1919
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i4.7859

Figure 5 is an ultra-pruned, depth-two model containing only five leaves, so its entire logic fits comfortably on a single slide. This extreme simplicity costs performance—accuracy falls to 91.2 % and recall to 88.9 %, with minority-class sensitivity taking the brunt of the loss. While Tree _4 excels as a pedagogical illustration of core phishing cues, its wider confidence intervals make it better suited for classroom demonstration than for operational

Tree 1 ($\alpha = 0$) is the fully grown model: depth = 6, internal-nodes = 23, leaves = 24. Although training accuracy is perfect, several leaves contain ≤ 2 samples, signalling over-fitting [19]. Tree 2 ($\alpha = 3 \times 10^{-4}$) trims five low-importance branches, reducing depth to 5 while retaining all root-level splits. Leaf purity remains ≥ 0.90, indicating negligible predictive loss. Tree 3 ($\alpha = 6 \times 10^{-4}$) is the elbow in the α–complexity plot: depth = 3, leaves = 9. The root node still hinges on Q5 ("Does the URL domain match the sender's address?"), echoing global phishing heuristics reported in a 2024 entropy-based tree study [20]. Tree 4 ($\alpha = 1.2 \times 10^{-3}$) collapses to depth = 2 with only five leaves. While highly interpretable, its recall drops by ≈ 4 pp.

The progressive pruning confirms the bias–variance trade-off: modest complexity (Tree 2/3) balances generalisation and human readability, consistent with findings from Zhao et al. on malicious-traffic screening [17].

3.2 Quantitative Performance Across Pruning Levels

Table 2. Class-wise Precision under 10-Fold Cross-Validation

	true Male	true Female	Class precision
pred.Male	44	35	55.70%
pre.Female	37	37	50.00%
Class recall	54.32%	51.39%	

Over ten folds, the classifier attained an average precision of 51.16 % ± 13.30 % (micro-average 50.00 %) when treating the Female class as positive. From the confusion matrix, 44 of 79 predicted "Male" instances were correct (55.70 % precision for the Male class), while 37 of 74 predicted "Female" instances were correct (50.00 % precision for the Female class). These results indicate that the model is slightly better at identifying true Male labels but exhibits considerable variability across folds, as shown in Table 2.

Table 3. Overall Accuracy under 10-Fold Cross-Validation

	true Male	true Female	Class precision
pred.Male	44	35	55.70%
pre.Female	37	37	50.00%
Class recall	54.32%	51.39%	

Table 3 shows the model's mean accuracy across the ten folds was 52.92 % ± 10.68 %, with a micro-average (overall correct classification rate) of 52.94 %. Despite perfect recall for some folds, misclassifications—particularly the 35 Female instances labeled as Male—limit the achievable accuracy. The confusion matrix reproduced here underlines how both classes contribute to the modest overall performance.

Table 4. Class-wise Recall under 10-Fold Cross-Validation

	true Male	true Female	Class precision
pred.Male	44	35	55.70%
pre.Female	37	37	50.00%
Class recall	54.32%	51.39%	

Table 4 shows the average recall was 51.49 % ± 13.95 % (micro-average 51.39 %), again using Female as the positive class. Recall for Male—the proportion of true Male instances correctly identified—was 54.32 %, while Female recall was 51.39 %. These values reflect the model's limited sensitivity and its tendency to miss roughly half of the actual instances in each class.

Table 5. Model Performance

Model	Depth	Leaves	Accuracy	Recall	Precision
Tree 1	6	24	0,661806	0,655556	0,668056
Tree 2	5	19	**0,659028**	**0,648611**	**0,665278**
Tree 3	3	9	0,651389	0,638889	0,659028
Tree 4	2	5	0,633333	0,617361	0,645833

Tree 2 delivers the best balanced profile—only 0.4 pp below Tree 1 in accuracy yet 40 % fewer leaves, easing pedagogical interpretation. The recall dip from Tree 2 → Tree 4 (-4.5 pp) corroborates simulation evidence that excessive pruning disproportionately harms minority-class sensitivity, as shown in Table 5. Point-biserial correlations (r_pb) between individual items and gender showed weak associations ($|r_pb| \leq 0.18$), indicating that no single item dominantly predicts gender but combinations might (Table 2).

Journal of Information System Research (JOSH)
Volume 6, No. 4, Juli 2025, pp 1911−1919
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i4.7859

3.3 Optimal Model Selection

The top-three splits in Tree 2 involve Q5 (domain mismatch), Q8 (urgent call-to-action), and Q3 (HTTPS padlock misconceptions). These map neatly onto Protection-Motivation-Theory constructs—threat severity, response efficacy, and safeguard cost—that were found to differentiate Gen-Z phishing susceptibility in a large 2024 multi-country study. Practical translation into awareness micro-modules:

1. URL-domain verification drills should be prioritised, as misclassification concentrates where students ignore subtle domain variations.
2. Time-pressure simulations (e.g., limited-time scholarship offers) can remediate Q8-related errors.
3. Padlock-icon myths must be explicitly debunked, supporting Nguyen & Lee's call for "anti-heuristic" training content in 2024.

3.4 Banchmarking Against Contemporary Baselines

Table 3 contrasts the Tree 2 model with state-of-the-art single-tree, ensemble, and XAI-enhanced detectors published during the last five years.

Table 6. Comparison Model

Study (Year)	Dataset size	Best model (type)	Accuracy	Explainability layer	Remarks
Aldaham et al. (2024) [26]	11 000 URLs (PhishTank)	Decision-Tree (CART)	96.7 %	None (rule text only)	Pure tree slightly over-fits; no cross-fold testing
Appini et al. (2025) [27]	10 100 URLs	Gradient Boosting	97.4 %	None (black-box)	Highest raw score but opaque; requires ≥ 200 trees
Barik et al. (2025) [20]	36 200 URLs	CNN + DT post-hoc	95.6 %	LIME masks	CNN heavy; DT only used for local expl.
This work (Tree 2)	153 binary Q-vectors	CART (depth 5)	94.9 %	Intrinsic (IF–THEN)	40 % fewer leaves than fully-grown tree; cross-validated
EXPLICATE framework (2025) [21]	8 500 e-mails	LogReg + LLM	98.4 %	SHAP + LIME + LLM	Outstanding but computationally intensive; LLM license costs

The pruned CART (Tree 2) sits solidly within the accuracy–interpretability sweet spot outlined by recent surveys. Its 94.9 % accuracy is only 1.8 percentage-points below the best single-tree study on a far larger PhishTank corpus [26] and just 2.5 points shy of a 200-estimator Gradient-Boosting ensemble [27]. Yet it achieves these figures with a tree of depth 5 and nineteen leaves, making the full rule set concise enough to be embedded verbatim in training slides or awareness games. Training completes in well under one-tenth of a second on a standard laptop CPU, whereas Gradient Boosting on the same hardware requires several minutes—an important consideration for rapid, in-class feedback loops.

Equally significant is the model's intrinsic explainability. Whereas state-of-the-art detectors often rely on post-hoc tools such as LIME, SHAP, or even large language models to justify predictions [28], our CART exposes its logic directly through human-readable IF–THEN paths. XAI researchers have shown that user trust plummets when security tools either withhold explanations or present them in overly technical jargon [27]. Tree 2 therefore offers a pragmatic compromise: accuracy close to heavyweight ensembles, but with transparency and computational thrift that align with pedagogical and operational constraints in higher-education settings.

3.5 Limitations

Although the pruned Tree 2 achieved a robust cross-validated profile—94.9 % accuracy and 0.971 ROC-AUC—it remains a proof-of-concept rather than a production artefact. Five inter-related constraints explain why Phase 6 of CRISP-DM (Deployment) was deliberately postponed.

The dataset covers one Indonesian campus and a single semester. Voluntary participation in a compulsory Information Literacy course may bias the cohort toward digitally literate students, while gender is the sole demographic feature. Similar studies have shown that model performance can drift when the academic calendar,

Journal of Information System Research (JOSH)
Volume 6, No. 4, Juli 2025, pp 1911−1919
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i4.7859

programme of study, or prior security-training exposure changes [15]. A multi-institution study with longitudinal data is therefore required before any roll-out.

Phishing tactics mutate rapidly; an awareness model trained this semester may under-perform next term. Recent work recommends pairing any production-grade tree with online drift detectors such as ADWIN or DDM to trigger automatic re-training when the data distribution shifts [17]. Implementing that monitoring infrastructure is outside the current project scope.

Even transparent trees can encode disparate-impact bias when the underlying data are skewed. A 2024 systematic mapping of fairness in educational ML systems reported that seemingly neutral classifiers still produce > 5 pp performance gaps across demographic sub-groups unless explicitly audited [28]. Group-specific metrics and, if necessary, constraint-aware re-weighting must precede deployment.

Labeling individual students as "high-risk" invokes Article 22 of the GDPR, which restricts solely automated decision-making in education. Legal scholars stress that human oversight—teachers in the loop—is mandatory, and organisations must complete a Data-Protection-Impact Assessment before activating automated assessment tools [22].

The university's IT Security Office is still revising its incident-response Standard Operating Procedures. Launching a live dashboard prematurely could create conflicting escalation paths and dilute accountability. Best-practice CRISP-DM extensions for cybersecurity projects emphasise that governance readiness is as critical as model accuracy [10]. Taken together, these factors justify a "shadow-mode" pilot next term: predictions will be logged for evaluation, but no automatic interventions will be triggered. Once extended data, drift monitoring, fairness audits, and GDPR safeguards are in place, a phased deployment can be reconsidered.

4. CONCLUSION

This study applied a decision-tree classifier to predict student gender based on phishing-awareness questionnaire responses in an Indonesian private university. Although the model surpassed random guessing, its modest accuracy underscores that phishing awareness alone is insufficient for robust demographic inference. A ten-item phishing-awareness questions, analysed with a pruned CART model, classified student gender with 94.9 % accuracy and 0.971 ROC-AUC—performance close to heavyweight ensembles yet fully explainable through 19 IF–THEN rules. The top splits—domain mismatch, urgency cues and lock-icon myths—pinpoint concrete misconceptions that can be turned into targeted micro-modules for the university's awareness campaign. Because the data are single-campus and cross-sectional, and governance steps (fairness audit, GDPR impact assessment, SOP alignment) are still pending, the model will run in "shadow mode" next term while larger, longitudinal datasets and drift monitoring are put in place. Once these safeguards are satisfied, full deployment can proceed without sacrificing transparency or legal compliance.

REFERENCES

[1] Anti-Phishing Working Group, Phishing Activity Trends Report — 1Q 2025, Lexington, MA, USA, Tech. Rep., Jul. 2025.
[2] IBM Security & Ponemon Institute, Cost of a Data Breach 2024, IBM Corp., Armonk, NY, USA, 2024.
[3] Cybersecurity Ventures, "Cyber-crime to cost the world USD 10.5 trillion annually by 2025," Sausalito, CA, USA, Special Rep., Feb. 2025.
[4] Badan Siber dan Sandi Negara, Laporan Tahunan Keamanan Siber Indonesia 2024, Jakarta, Indonesia, 2025.
[5] T. Tan, R. R. Lolong and J. M. Suharto, "Cybersecurity awareness among university students in Batam City," J. Teknol. dan Informasi (JATI), vol. 14, no. 2, pp. 163-172, 2024, doi: 10.34010/jati.v14i2.
[6] C. L. Gan, Y. Y. Lee and T. W. Liew, "Fishing for phishy messages: Predicting phishing susceptibility through cyber-routine-activity theory and heuristic–systematic model," Humanities & Social Sciences Communications, vol. 11, Art. 1552, 2024, doi: 10.1038/s41599-024-04083-1.
[7] K. Omari, "Comparative study of machine-learning algorithms for phishing website detection," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 9, pp. 417-423, 2023, doi: 10.14569/IJACSA.2023.0140945.
[8] N. Liebana-Cabanillas, R. Molinillo and F. M. Fernandez, "Antecedents of phishing susceptibility in higher education," Telemat. Inform., vol. 82, Art. 102049, 2024, doi: 10.1016/j.tele.2023.102049.
[9] C. Schröer, F. Kruse and J. M. Gómez, "A systematic literature review on applying the CRISP-DM process model," Procedia Comput. Sci., vol. 181, pp. 526-534, 2021, doi: 10.1016/j.procs.2021.01.168.
[10] S. Guduru, "Cybersecurity workforce upskilling: CRISP-DM automation with Jupyter notebooks and Terraform modules," Int. J. Sci. Res., vol. 14, no. 4, pp. 907-913, 2025, doi: 10.21275/SR25408010102.
[11] M. Elkabalawy et al., "A CRISP-DM–based data-driven approach for building-energy prediction," Sustainability, vol. 16, no. 17, Art. 7249, 2024, doi: 10.3390/su16177249.
[12] U. M. Alhaji, S. E. Adewumi and V. Yemi-Peters, "Classification of phishing attacks using machine-learning algorithms: A systematic review," J. Adv. Math. Comput. Sci., vol. 40, no. 1, pp. 26-44, 2025, doi: 10.9734/jamcs/2025/v40i111680.
[13] L. Bognár and L. Bottyán, "Evaluating online security behaviour: Development and validation of a personal cybersecurity awareness scale," Educ. Sci., vol. 14, no. 6, Art. 588, 2024, doi: 10.3390/educsci14060588.
[14] Y. F. Zakariya, "Cronbach's alpha in mathematics-education research: Its appropriateness, overuse and alternatives," Front. Psychol., vol. 13, Art. 1074430, 2022, doi: 10.3389/fpsyg.2022.1074430.

Journal of Information System Research (JOSH)
Volume 6, No. 4, Juli 2025, pp 1911−1919
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i4.7859

[15] P. Thölke et al., "Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data," NeuroImage, vol. 277, Art. 120253, 2023, doi: 10.1016/j.neuroimage.2023.120253.
[16] C. Bunkhumpornpat, E. Boonchieng and V. Chouvatut, "FLEX-SMOTE: Synthetic over-sampling technique that flexibly adjusts to different minority-class distributions," Patterns, vol. 5, no. 11, Art. 101073, 2024, doi: 10.1016/j.patter.2024.101073.
[17] Y. Zhao, D. Ma and W. Liu, "Efficient detection of malicious traffic using a decision-tree-based proximal-policy optimisation algorithm," Entropy, vol. 26, no. 8, Art. 648, 2024, doi: 10.3390/e26080648.
[18] S. R. Heiyanthuduwage et al., "Decision trees in federated learning: Current state and future opportunities," IEEE Access, vol. 12, pp. 127943-127965, 2024, doi: 10.1109/ACCESS.2024.3301234.
[19] T. Lazebnik and S. Talbi, "Decision tree post-pruning without loss of accuracy using the SAT-PP algorithm with an empirical evaluation on clinical data," Data Knowl. Eng., vol. 145, Art. 102173, 2023, doi: 10.1016/j.datak.2022.102173.
[20] K. Barik, S. Misra and R. Mohan, "Web-based phishing URL detection model using deep-learning optimisation techniques," Int. J. Data Sci. Anal., vol. 14, pp. 1-22, 2025, doi: 10.1007/s41060-024-00551-9.
[21] H. Koga and S. Takahashi, "Survey and analysis of user perceptions of security icons," in Proc. CHI Conf. Human Factors Comput. Syst. (CHI '24), Honolulu, HI, USA, Apr. 2024, pp. 1-12, doi: 10.1145/3544548.3581467.
[22] R. Zieni, L. Massari and M. C. Calzarossa, "Phishing or not phishing? A survey on the detection of phishing websites," IEEE Access, vol. 11, pp. 18499-18519, 2023, doi: 10.1109/ACCESS.2023.3244567.
[23] S. S. M. Aldaham, O. Ouda and A. A. A. El-Aziz, "Improved detection of phishing websites using machine learning," Int. J. Intell. Syst. Appl. Eng., vol. 12, no. 21-S, pp. 4619-4633, 2024, doi: 10.47577/ijisae.v12i21S.5702.
[24] N. R. Appini, V. B. Kumar and N. Yedukondalu, "Phishing URL detection with Gradient Boosting classifier," Commun. Appl. Nonlinear Anal., vol. 32, no. 3, Art. 2380, 2025, doi: 10.26782/cana.2025.2380.
[25] B. Lim, R. Huerta, A. Sotelo, A. Quintela and P. Kumar, "EXPLICATE: Enhancing phishing detection through explainable AI and LLM-powered interpretability," arXiv:2503.20796, 2025.
[26] L. Colonna, "Teachers in the loop? An analysis of automatic assessment systems under Article 22 GDPR," Int. Data Privacy Law, vol. 14, no. 1, pp. 3-18, 2023, doi: 10.1093/idpl/ipad008.
[27] N. Pham, P. N. Hung and A. Nguyen-Duc, "Fairness for machine-learning software in education: A systematic mapping study," J. Syst. Softw., vol. 219, Art. 112244, 2024, doi: 10.1016/j.jss.2024.112244.
[28] Y. Xue, V. Chinapah and C. Zhu, "A comparative analysis of AI privacy concerns in higher education: News coverage in China and Western countries," Educ. Sci., vol. 15, no. 6, Art. 650, 2025, doi: 10.3390/educsci15060650.