Evaluating Cybersecurity Awareness in Employees Using Gameplay: Data and Machine Learning Models

Mike Wa Nkongolo[1] and Mahmut Tokmak[2]
[1]University of Pretoria, Department of Informatics, South Africa
[2]Burdur Mehmet Akif Ersoy University, Department of Management Information Systems, Türkiye

mike.wankongolo@up.ac.za
mahmuttokmak@mehmetakif.edu.tr

Abstract: Cyber-attacks continue to pose persistent challenges within professional environments. Human error remains a critical vulnerability, frequently leading to security breaches through credential misuse and social engineering tactics. Traditional cybersecurity training approaches often lack effectiveness when not adapted to the dynamic threat landscape. This study presents *CyberEmployee*, a serious game developed to enhance cybersecurity awareness among employees through interactive learning. The objective is to assess employees' awareness levels by analysing gameplay data using machine learning techniques. Data were collected via the game's integrated scoreboard, which tracked user behaviors and performance patterns. The resulting dataset was analysed using multiple machine learning algorithms, including Random Forest, Support Vector Machines (SVM), XGBoost, K-Nearest Neighbors (KNN), and Logistic Regression. Experimental results demonstrated accuracy rates ranging from 86% to 100% and F1-scores from 75% to 100%. The highest performance—100% accuracy and 100% F1-score—was achieved using the Random Forest and XGBoost models. This analysis indicates that ensemble learning methods outperform other classifiers in employee classification. Furthermore, gameplay duration and player score were identified as key predictive features. These findings indicate the potential of serious games combined with machine learning for data-driven cybersecurity training frameworks.

Keywords: Cybersecurity awareness, Cybersecurity training, Gamification, Machine learning

1. Introduction

Cybersecurity threats are becoming complex, posing significant challenges to organisations across various sectors. Despite advancements in technical defenses, human error remains a leading cause of security breaches, often due to social engineering, credential misuse, and poor security awareness (Arif et al., 2025). Traditional cybersecurity training approaches—based on static content such as presentations or videos—have proven insufficient in engaging users and adapting to the evolving threat landscape. These methods lack interactivity, fail to simulate realistic attack scenarios, and often do not promote the cognitive and behavioral changes needed to prevent breaches. This research investigates the effectiveness of modern cybersecurity training methodologies, with a specific focus on serious games as tools for enhancing users' cybersecurity preparedness. The study focuses on the evaluation of *CyberEmployee*, an interactive serious game developed to address the limitations of traditional cybersecurity training approaches. In response to the growing demand for more effective cybersecurity training, organisations are adopting human-centric approaches that emphasize user engagement, behavior modification, and proactive awareness (Deibert, 2018). One such approach is CyberEmployee, a training game designed to enhance employees' ability to detect threats and respond to incidents through simulated cyber scenarios (Nkongolo, 2024). CyberEmployee supports the development of long-term cybersecurity competencies by promoting situational awareness in realistic contexts. Although human-centric training methods show considerable promise, their practical implementation presents several challenges. Organisational readiness, compatibility with existing IT infrastructure, resource constraints, and the need for active stakeholder involvement are critical factors that can influence the success of such training. Growing interest has emerged in the application of various learning theories for cybersecurity training, primarily due to their capacity to provide interactive and engaging learning experiences. These theories form the foundation for designing educational games that enhance user comprehension. These approaches help learners actively construct knowledge through problem-solving, and real-time decision-making. As demonstrated in recent studies, the integration of learning theories not only improves cybersecurity awareness but also supports behavior change and skill acquisition, making them essential for developing effective human-centric training programs. Chattopadhyay, Maschinot, and Nestor (2021) analyse several popular cybersecurity educational games and *Capture the Flag* (CTF) platforms using four key benchmarks: the CSEC2017 guidelines, the Cybersecurity Assessment Tools (CATS), NSA GenCyber concepts, and the NICE framework. Their findings illustrate how each game aligns with curricular domains, core cybersecurity concepts, and specialized skill sets. This analysis offers guidance for educators and practitioners, enabling informed selection of games and supporting gap analysis for curriculum development and training strategies. While the study do not state a specific learning theory, it implicitly supports *constructivisim*. Tempestini et al. (2024) explored a gamified

training program aimed at increasing cybersecurity awareness among young adults. The study is grounded in *cognitive learning* theory, emphasizing mental processes such as problem-solving and decision-making. The tactical gaming approach proved effective in enhancing participants' understanding of cybersecurity concepts. Nevertheless, the study faced limitations, including a small sample size and broad content coverage, which may have diluted the effectiveness in addressing specific cybersecurity topics. In light of the limitations of traditional cybersecurity training methods and the growing need for interactive learning tools, this study explores the potential of CyberEmployee in enhancing cybersecurity awareness. The research is guided by the following question: *How can employee cybersecurity awareness be effectively evaluated through gameplay data using machine learning models?*

2. Related Work

Tokmak (2023) evaluated cybersecurity awareness among students using machine learning. After applying descriptive statistics, several models were used to classify awareness levels as low, moderate, or high. Most students showed moderate awareness, with limited knowledge of phishing attacks. Gender differences were minimal, though female students showed greater concern for data integrity. IT students displayed higher awareness due to their coursework. Multilayer Perceptron (MLP) and Support Vector Machine (SVM) achieved the best classification performance, demonstrating the effectiveness of machine learning in assessing cybersecurity awareness. Rismayanti (2024) conducted a study to predict online gaming behavior using machine learning techniques. A dataset from *Kaggle* containing player demographics and in-game metrics is employed to forecast player engagement levels. The study utilised the Gaussian Naïve Bayes (GNB) algorithm within a supervised learning framework and evaluated model performance using accuracy as the primary metric. Results demonstrated the GNB effectiveness in predicting player engagement, but the study acknowledged limitations related to resource constraints, including limited access to technology and lack of technical expertise. Kuna (2024) explored the application of machine learning techniques to predict the popularity of video games by analysing patterns in player data. The study employed the K-Nearest Neighbors (KNN) algorithm to forecast player trends, aiming to improve user experience (UX). Although it followed a supervised learning approach, the evaluation metrics were not specified. Smerdov et al. (2023) investigated an AI-enabled approach to predicting *eSports* player performance using data collected from sensors. The study employed Convolutional Neural Networks (CNNs) with regression metrics used for evaluation. The results indicated strong potential for using sensor data to forecast in-game player performance. Nevertheless, it lacked detailed information on specific algorithms used, as well as comprehensive evaluation metrics, which limits the ability to fully assess the model's reproducibility (Table 1). Syed et al. (2020) developed a video recognition system based on a CNN model to digitize the *blackjack* game by detecting players in real-time, with the goal of constructing accurate player profiles. The system, rooted in supervised learning, achieved approximately 97% accuracy. Despite these promising results, the difficulty of replicating the variability of real-world settings within a controlled environment impacted the model's performance. Table 1 shows a comparative analysis of various studies that explore the integration of machine learning and gamified approaches in cybersecurity training.

Table 1: Comparative analysis of studies on gaming and cybersecurity training

Author	Algorithm	Metrics	Results	Limitations
Syed et al. (2020)	CNN	Accuracy	95% Accuracy	May not reflect real-world variability
Smerdov et al. (2023)	CNNs	Regression	Strong prediction	Lack of evaluation results
Tokmak (2023)	SVM, MLP	Accuracy	98% MLP Accuracy and 95% SVM Accuracy	Dataset limited to university students
Kuna (2024)	KNNs	Not specify	UX understanding	Limited comparative analysis
Rismayanti (2024)	GNB	Accuracy	Effective player prediction	Ressource constraints
Nkongolo, Sithole, and Sewnath (2025)	Random Forest	F1, ROC	100 % ROC AUC	Replayability
This Study	XGBoost, SVM, Random Forest	Accuracy, Precision, Recall, F1, ROC	100% XGBoost Accuracy	Limited generalizability

The most widely applied approach in gaming analytics is machine learning, which has been employed in numerous studies to predict player engagement, analyse game popularity, and assess in-game performance (Table 1). Some of these studies utilise sensor data to evaluate player behavior, while others rely on demographic and in-game metrics to classify engagement levels. For instance, *DotaGame* demonstrates the effectiveness of machine learning models in outcome prediction, though its findings may not be generalizable (Akhmedov and Phan, 2021). Similarly, the controlled experimental conditions used in Ghazali et al. (2023) ensured high accuracy but limited the ability to capture the variability of real-world contexts. Each study employed different methodologies, algorithms, and evaluation metrics—some achieving high accuracy using deep learning techniques, while others lacked sufficient methodological transparency, making direct comparisons difficult (Table 1). This study builds on earlier work by Nkongolo, Sithole, and Sewnath (2025), which showed that gamification effectively improved students' cybersecurity awareness. The current study extends Nkongolo, Sithole, and Sewnath (2025)'s research by focusing on employees rather than students, offering more practical insights into how the game can enhance cybersecurity skills in real workplace settings. Rather than examining isolated gameplay metrics in controlled environments, the research analyses employees actions within dynamic and uncontrolled real-world scenarios. The proposed approach incorporates a wider range of gaming data and employee behavior indicators, making it more applicable to machine learning analysis (Table 1).

3. CyberEmployee Gameplay

The CyberEmployee game is played over 13 rounds on a virtual board (Figure 1). Let $B = \{b_1, b_2, b_3, \dots, b_{13}\}$ represent the sequence of rounds in the game. In each round b_t (where t ranges from 1 to 13), both players make a move: $m_t^A \in A$ denotes the attacker's chosen token at round t, and $m_t^D \in D$ represents the defender's response token at round t. Each round is represented as a pair: $b_t = (m_t^A, m_t^D)$. And the judging agent (J) evaluates the effectiveness of each move pair using a reward function (1).

$$J(m_t^A, m_t^D) \rightarrow (r^A, r^D) \quad (1)$$

where:

$r^A \in \{0,1\}$ is the reward for the attacker and $r^D \in \{0,1\}$ is the reward for the defender. These rewards are assigned as (1, 0) if the attacker's move is successful and the defender fails to counter. (0, 1) if the defender's move successfully counters the attacker, and (0, 0) if neither player selects an optimal move (Table 2). The total score for each player is computed by summing their respective rewards across all rounds (2):

$$R^A = \sum_t J(m_t^A, m_t^D)_1 \quad (2)$$

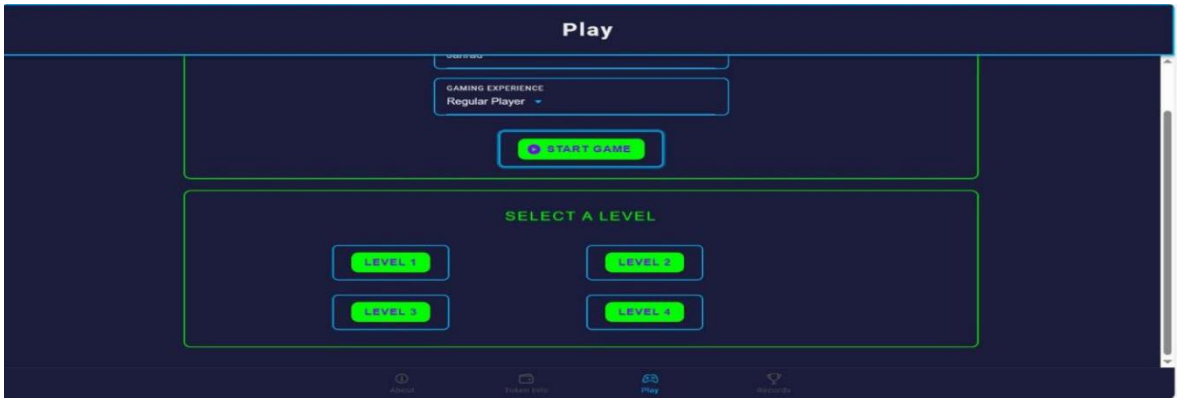
$$R^D = \sum_t J(m_t^A, m_t^D)_2 \quad (2.1)$$

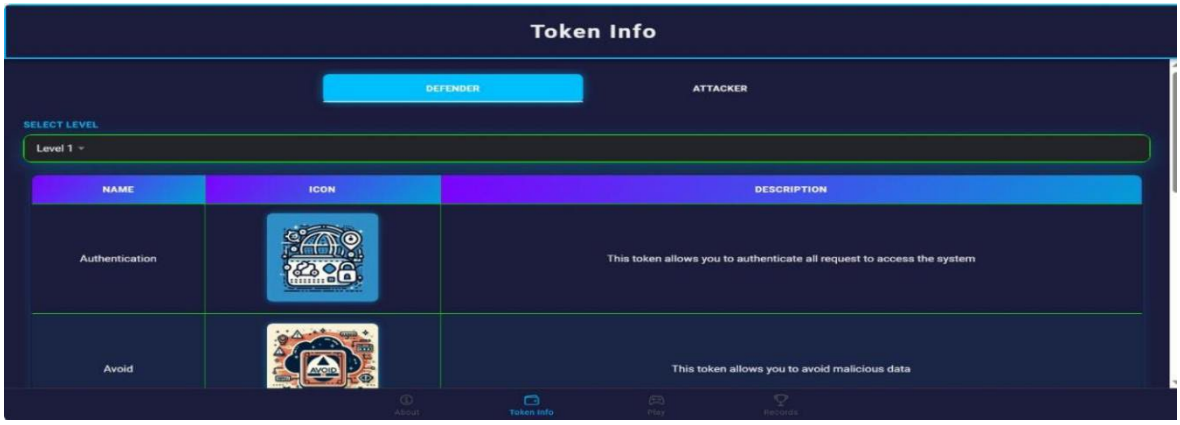
where $J(m_t^A, m_t^D)_1$ and $J(m_t^A, m_t^D)_2$ denote the attacker's and defender's rewards respectively for round t.

Table 2: Token pair scoring matrix

Attacker Token	Attacker Strategy	Defender Token	Defender Strategy	Score (A:D)
A1 - Email	Malicious e-mail	D1 - Denying	Block/deny access	0 : 1
A2 - Phone	Malicious phone call	D2 - Identification	Verify identity	0 : 1
A3 - Password	Stolen password	D3 - Upload	Upload controls	1 : 0
A4 - Click	Phishing link	D4 - Avoid clicking	Refuse to click unknown links	0 : 1

The game ends when all 13 pairs of tokens have been placed on the board. The final outcome is determined as $R^A > R^D$ if the attacker wins, $R^D > R^A$ if the defender wins, and $R^A = R^D$ if the game is a draw (Figure 1).


Play
GAMING EXPERIENCE
Regular Player
START GAME
SELECT A LEVEL
LEVEL 1
LEVEL 2
LEVEL 3
LEVEL 4


Token Info
DEFENDER
ATTACKER
SELECT LEVEL
Level 1
NAME
ICON
DESCRIPTION
Authentication
This token allows you to authenticate all request to access the system
Avoid
This token allows you to avoid malicious data


Token Info
DEFENDER
ATTACKER
SELECT LEVEL
Level 1
NAME
ICON
DESCRIPTION
Data
This token represents malicious data
Email
This token represents malicious email

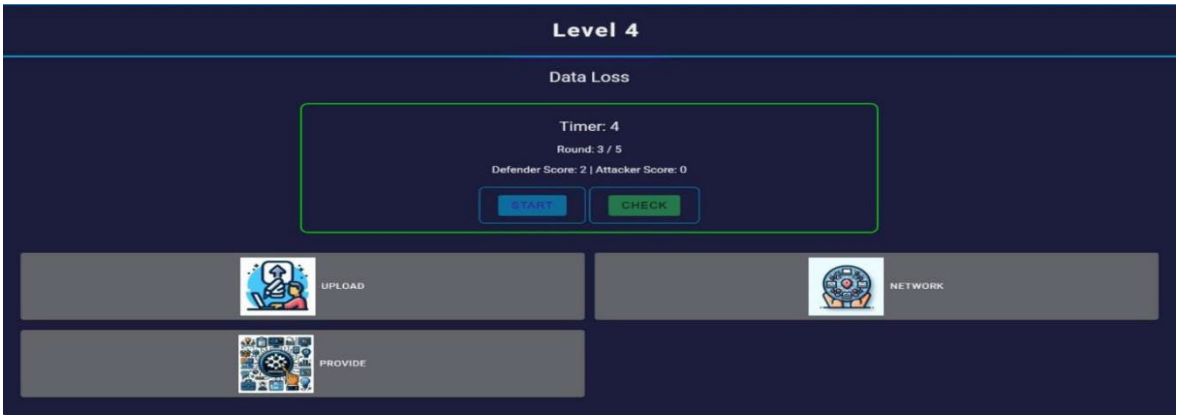

Level 4
Data Loss
Timer: 4
Round: 3 / 5
Defender Score: 2 | Attacker Score: 0
START
CHECK
UPLOAD
NETWORK
PROVIDE

Figure 1: Overview of CyberEmployee level selection interface, token information, gameplay snapshot (Level 4, Round 3 of 5), and corresponding scoreboard records

Each player chooses a strategy that determines the sequence in which they place their tokens. Let $\sum^A$ be the set of all permutations of attacker tokens and $\sum^D$ be the set of all permutations of defender tokens (Figure 1). A strategy is defined as a function $\sigma^A: \{1,2,\ldots,13\} \rightarrow A$ or $\sigma^D: \{1,2,\ldots,13\} \rightarrow D$ where each token is used exactly once. An optimal strategy σ^* maximizes a player's expected cumulative score against an opponent's strategy. Each attacker strategy (A_i) has an ideal defending strategy (D_j) mapped based on common cybersecurity best practices. The CyberEmployee game unfolds over 4 levels, each denoted L_i for $i = 1, 2, 3, 4$, with each level consisting of 5 rounds, totaling 20 rounds (Figure 2). Each round is determined by a reward function $J(m_t^A, m_t^D) \rightarrow (r^A, r^D) \in \{0,1\}\, x\, \{0,1\}$. As levels progress, the number of effective (i.e., winning) defender tokens decreases while attacker options increase (Figure 2). Let $W^{D(i)}$ denote the number of winning defender tokens and $W^{A(i)}$ the number of winning attacker tokens at level L_i. Then, the transition obeys (2.2).

$$W^{D(1)} > W^{D(2)} > W^{D(3)} > W^{D(4)} >, and\ W^{A(1)} < W^{A(2)} < W^{A(3)} < W^{A(4)} < W^{A(i)} \quad (2,2)$$

This reflects a growing asymmetry, where $|D_i|$ decreases and $|A_i|$ increases, adding pressure on the defender (Figure 2). Each strategy is a bijection $\sigma^A: \{1,2,\ldots,20\} \rightarrow A$ and $\sigma^D: \{1,2,\ldots,20\} \rightarrow D$, where players select tokens without repetition. The reduction in $W^{D(i)}$ across levels reduces the defender's available counter-strategies, making optimal play increasingly difficult.

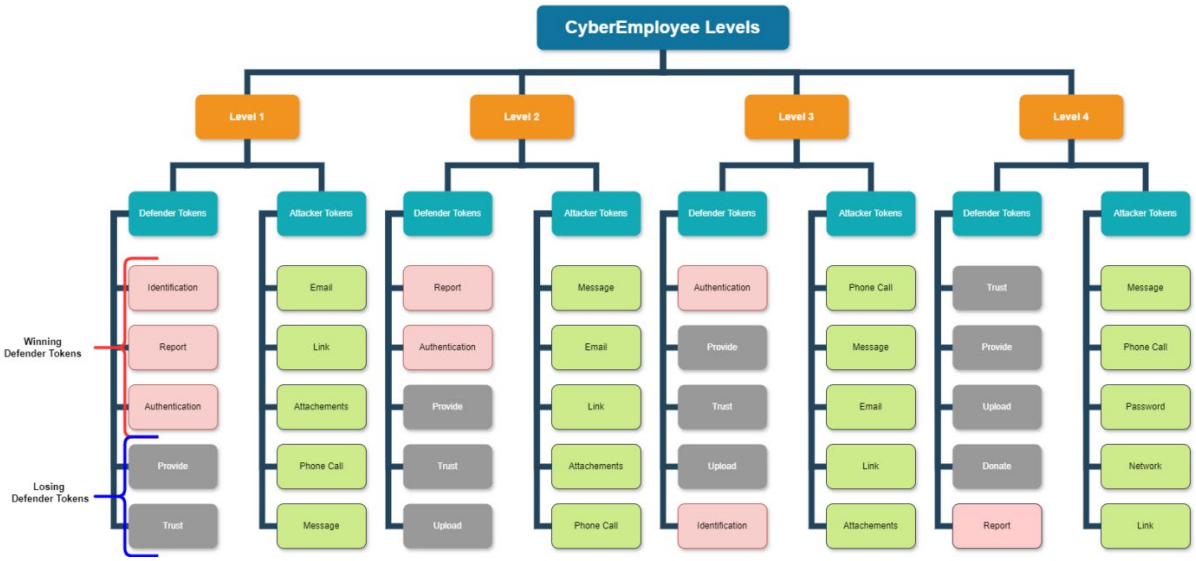


Figure 2: CyberEmployee Levels

4. Methodology

The study frameworks are illustrated in Figure 3 and Figure 4. The process begins with the data collection phase, in which employees play the CyberEmployee game to generate data (Figure 3). Participants from an IT organisation voluntarily engaged in gameplay after signing informed consent forms. Ethical clearance was obtained from the affiliated institution prior to data collection.

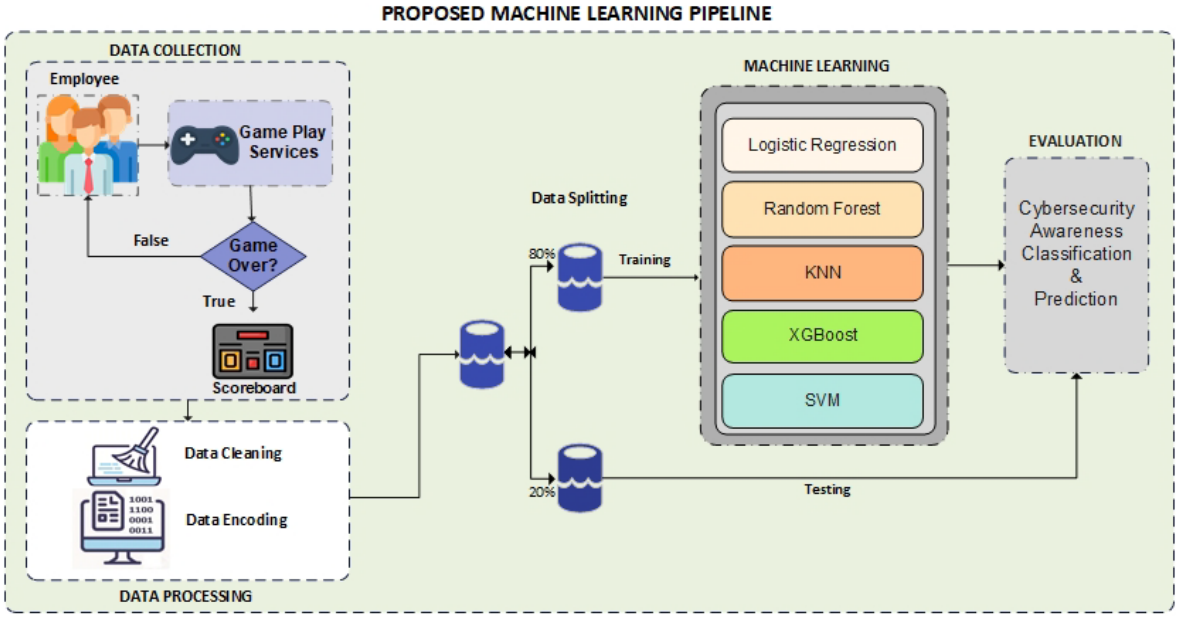


Figure 3: The proposed machine learning pipeline

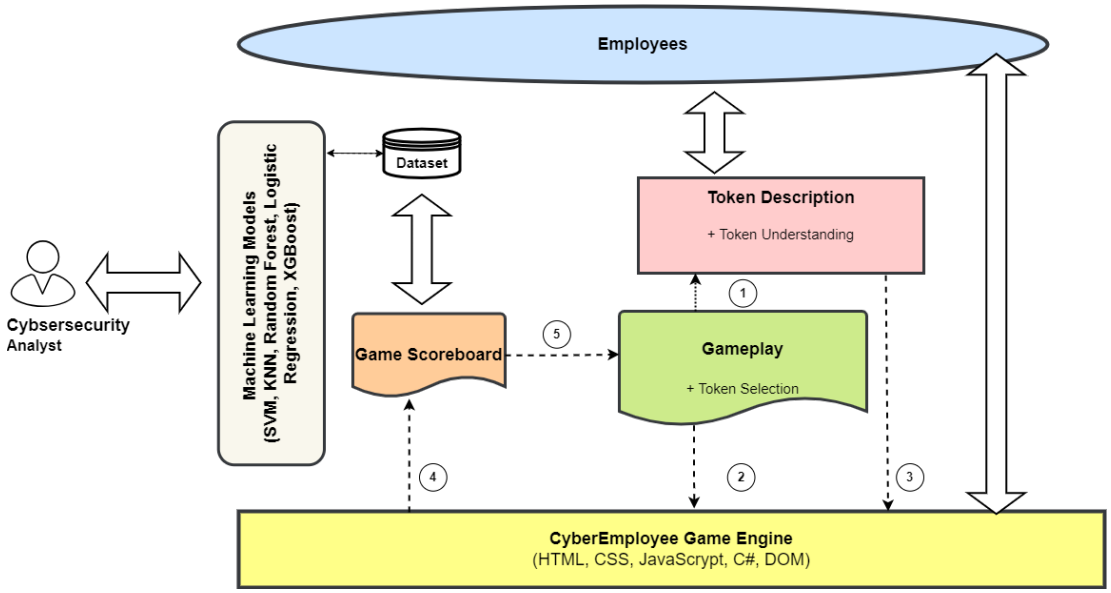


Figure 4: Data collection steps

Upon completion of the game, participant performance levels were automatically classified into beginner, intermediate, and expert tiers, as recorded on the CyberEmployee scoreboard (Figure 1). This data is exported in CSV format for further processing (Figure 4). The second phase involved data pre-processing to ensure data quality. In the third phase, the dataset is encoded using *One-Hot Encoding* to convert categorical variables into a machine-learning-friendly format (Figure 4). The resulting dataset is then split into training (80%) and testing (20%) subsets (Figure 3). Five supervised machine learning algorithms—Random Forest (RF), Support Vector Machines (SVM), XGBoost, K-Nearest Neighbors (KNN), and Logistic Regression (LR) (Figure 4)—were trained and evaluated to classify and predict employee cybersecurity awareness levels (Horvat and Job, 2020). Finally, the performance evaluation phase compared the models using metrics such as accuracy, precision, recall, F1

score, and receiver operating characteristic (ROC) curves (Horvat and Job, 2020), with the goal of identifying the most effective algorithm.

5. Results

In this experiment, the defender (employee) competed against the attacker (a computer program). A comparative analysis of machine learning models revealed that the Random Forest classifier accurately identified 56 instances as defenders patterns and classified only 1 gameplay as a draw (Figure 5). Moreover, this algorithm effectively predicted cybersecurity skill levels, classifying 53 employees as beginners, 32 as experts, and 15 as intermediate (Figure 5). These results suggest that employees in this organisation demonstrate strong cybersecurity awareness. In contrast, other algorithms—such as KNN exhibited higher rates of misclassification in both defender identification and skill level prediction (Figure 6). Figure 7 illustrates the key features influencing the model's predictions. The game outcome (Winner) and the employee identifier (Nickname) exhibit minimal impact on the classification results. In contrast, the most influential factors are the employee's performance score (DefenderScore), the duration of gameplay (Time (sec)), and the computer's performance (AttackerScore). These features play a central role in shaping the model's ability to accurately predict cybersecurity skill levels of the employees (Table 3 and Figure 8). This suggests that the game is capturing meaningful behavioral signals that reflect employee's decision-making quality and situational awareness during gameplay. The low importance of superficial features like Outcome (Winner) and Identifier (Nickname) reinforces that the model is not biased by outcomes or identity, but by in-game behavioral metrics (Figure 7).

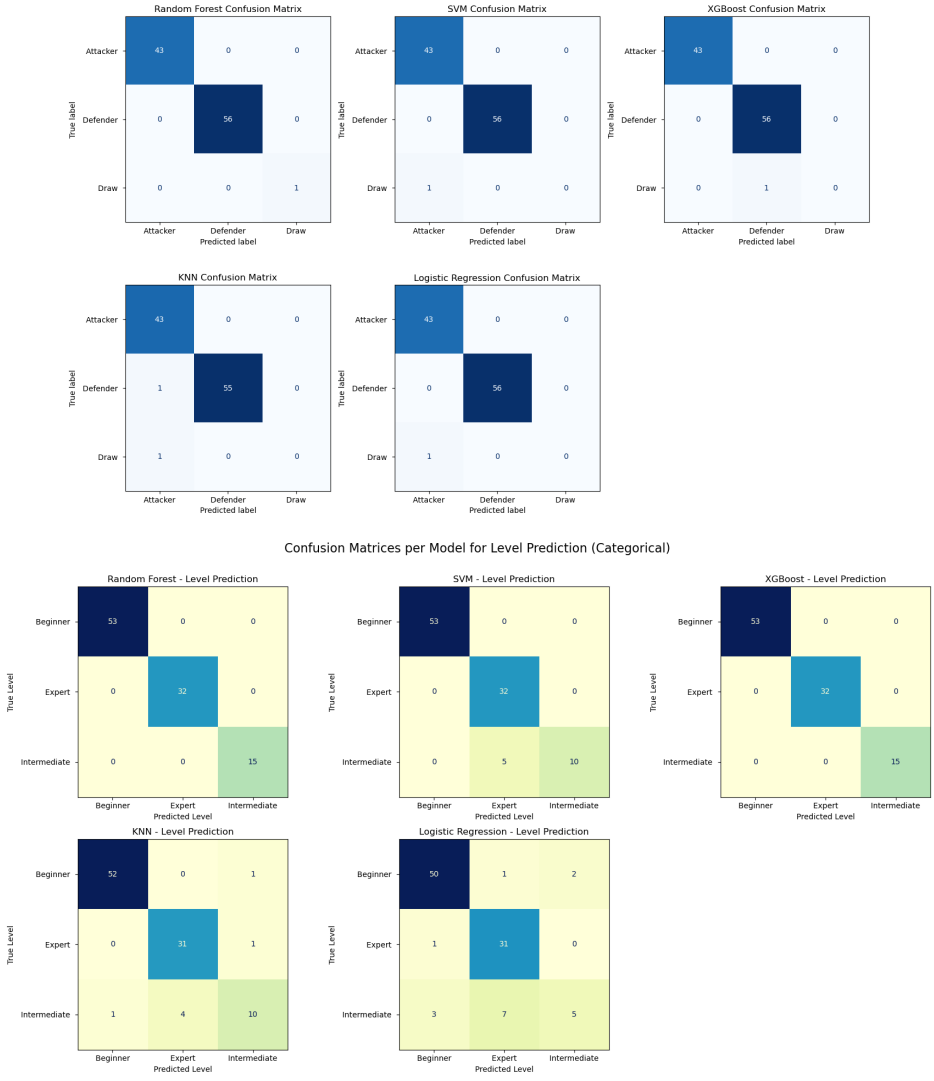


Figure 5: Overall machine learning results

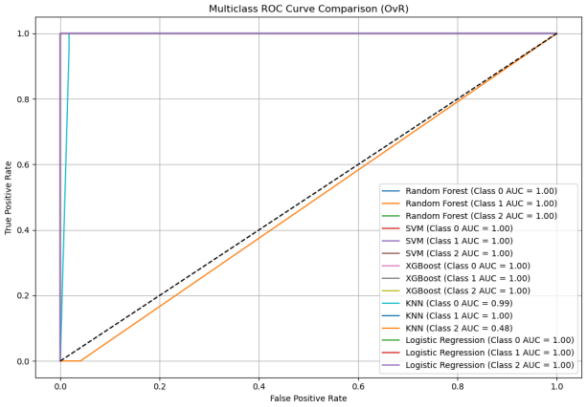


Figure 6: The ROC curve at various threshold settings

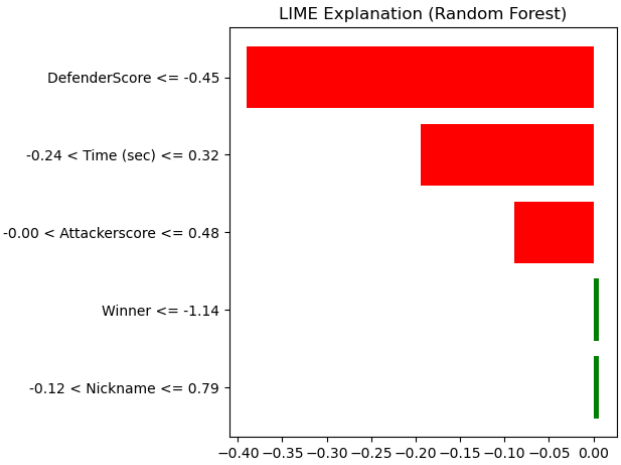


Figure 7: Features influencing the model's prediction

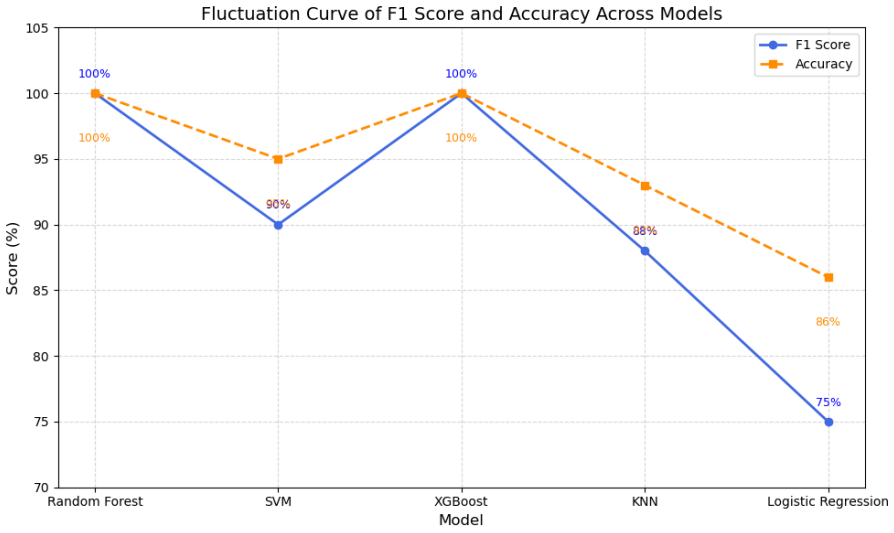


Figure 8: Each model's performance

Features such as DefenderScore, Time, and AttackerScore are the most influential, reflecting genuine cognitive and strategic engagement during gameplay. This supports the game's role as both an assessment and training tool, validating CyberEmployee as a practical approach to reducing human-centric cybersecurity risks.

Table 3: Comparative analysis of machine learning models

Model	Accuracy	Precision	Recall	F1 score	ROC AUC
Random Forest	100%	100%	100%	100%	100%
SVM	95%	95%	88%	90%	99%
XGBoost	100%	100%	100%	100%	100%
KNN	93%	90%	87%	88%	98%
Logistic Regression	86%	81%	74%	75%	97%

Figure 9 illustrates the overall distribution of participant responses across all cybersecurity awareness categories in the CyberEmployee game survey. The chart highlights that a majority of participants responded with "Strongly Agree" and "Agree", together accounting for over 80% of total responses. This positive feedback demonstrates the effectiveness of CyberEmployee in promoting key human-centric cybersecurity skills such as threat awareness, problem-solving, and self-reflection (Figure 10). Conversely, negative responses such as "Disagree" and "Strongly Disagree" were minimal, reinforcing the game's positive impact on employee cybersecurity awareness.

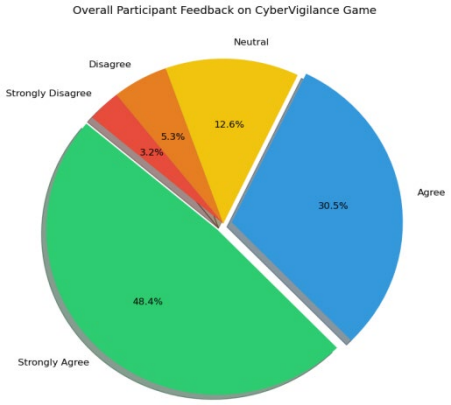


Figure 9: Overall survey response

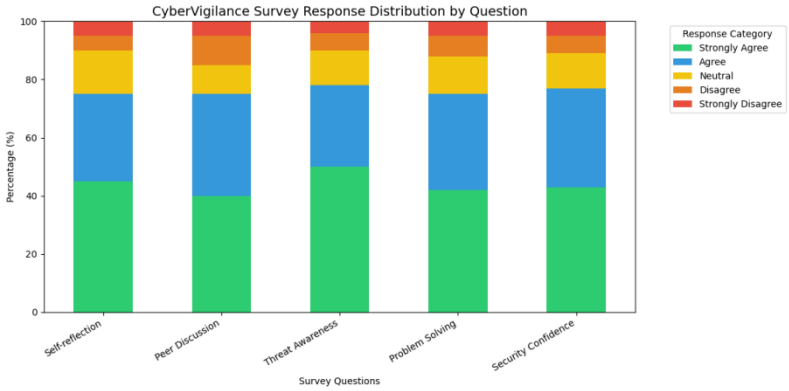


Figure 10: Survey response by question

6. Conclusion

Human error remains a leading cause of cybersecurity breaches, yet traditional cybersecurity training often fails to adequately prepare employees for the evolving threat landscape. To address this gap, the present study introduces CyberEmployee, a digital card game designed to evaluate and enhance employee awareness in realistic cybersecurity scenarios. Findings from the study indicate that the game offers an interactive learning environment that replicates real-world attack dynamics. The effectiveness of CyberEmployee is further supported by post-game survey responses. Employees reported a notable improvement in their understanding of cybersecurity risks, with many highlighting how the game encouraged critical reflection and peer discussion. In conclusion, the study demonstrates that gamification is a powerful tool for advancing cybersecurity awareness

through experiential learning. The research makes a strong case for shifting from passive cybersecurity training to dynamic, and scenario-based models. Future work will explore adaptive gameplay tailored to individual risk profiles.

Ethics declaration: Ethical approval was duly obtained prior to the commencement of the study. The dataset used is publicly available at: https://www.kaggle.com/dsv/9480976

AI declaration: Artificial Intelligence (AI) tools were used to assist in generating illustrative tokens. Bing AI was used for assistance in this regard: https://www.bing.com/chat. The use of AI was supplementary and did not replace original scholarly work or analysis.

References

Akhmedov, K. and Phan, A.H., (2021). Machine learning models for DOTA 2 outcomes prediction. *arXiv preprint arXiv:2106.01782. https://doi.org/10.48550/arXiv.2106.01782*

Arif, M., Badila, M., Warden, J. M., & Ur Rehman, A. (2025). A study of human factors toward compliance with organization's information security policy. Information Security Journal: A Global Perspective, 34(3), 235–250. https://doi.org/10.1080/19393555.2025.2457702

Chattopadhyay, A., Maschinot, C., & Nestor, L. (2021). Mirror mirror on the wall — What are cybersecurity educational games offering overall: A research study and gap analysis. In 2021 IEEE Frontiers in Education Conference (FIE) (pp. 1–8). Lincoln, NE, USA: IEEE. https://doi.org/10.1109/FIE49875.2021.9637224

Deibert, R. J. (2018). Toward a human-centric approach to cybersecurity. Ethics & International Affairs, 32(4), 411–424. https://doi.org/10.1017/S0892679418000618

Ghazali, E.M., Al Halbusi, H., Abdel Fattah, F.A.M., Hossain Uzir, M.U., Mutum, D.S. and Tan, F.-L. (2023), "A study of player behavior and motivation to purchase Dota 2 virtual in game items", Kybernetes, Vol. 52 No. 6, pp. 1937-1961. https://doi.org/10.1108/K-08-2021-0678

Horvat, T., & Job, J. (2020). The use of machine learning in sport outcome prediction: A review. WIREs Data Mining and Knowledge Discovery, 10(1), e1380. https://doi.org/10.1002/widm.1380

Kuna, V. (2024). Machine learning models in predicting gaming popularity: A comparative analysis. Multidisciplinary Research, 6(3), 1-7.

Nkongolo, M. W. (2024). Gamified security tactics through digital card game models. In Proceedings of the 2024 4th International Multidisciplinary Information Technology and Engineering Conference (IMITEC) (pp. 48–55). IEEE. Vanderbijlpark, South Africa, https://doi.org/10.1109/IMITEC60221.2024.10850994

Nkongolo, M. W., Sithole, T., & Sewnath, J. (2025, March). Cybersecurity awareness through interactive learning using the CyberVigilance game. In Proceedings of the 20th International Conference on Cyber Warfare and Security (pp. 501–510). Williamsburg, Virginia, USA: Academic Conferences International Limited. https://doi.org/10.34190/iccws.20.1.3207

Rismayanti, N. (2024). Predicting Online Gaming Behaviour Using Machine Learning Techniques. Indonesian Journal of Data and Science, 5(2), 133-143. https://doi.org/10.56705/ijodas.v5i2.166

Smerdov, A., Somov, A., Burnaev, E., & others. (2023). AI-enabled prediction of video game player performance using the data from heterogeneous sensors. Multimedia Tools and Applications, 82(11021–11046). https://doi.org/10.1007/s11042-022-13464-0

Syed, D., Gandhi, N., Arora, A., & Kadam, N. (2020). DeepGamble: Towards unlocking real-time player intelligence using multi-layer instance segmentation and attribute detection. In Proceedings of the 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 376–383). IEEE. https://doi.org/10.1109/ICMLA51294.2020.00067

Tempestini, G., Merà, S., Palange, M. P., Bucciarelli, A., & Di Nocera, F. (2024). Improving the cybersecurity awareness of young adults through a game-based informal learning strategy. Information, 15(10), 607. https://doi.org/10.3390/info15100607

Tokmak, M. (2023). Determination of Students' Cyber Security Awareness Levels with Machine Learning Methods. Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 28(2), 451–466. https://doi.org/10.53433/yyufbed.1181694