JIT Mahasarakham University, Thailand
http://jit.it.msu.ac.th

Journal of Applied Informatics and Technology, 7(2), 375-388.

Research Article

Utilizing Association Rule Mining to Understand Phishing Risk Awareness Levels of Thai University Academic Staff

Pita Jarupunphol[1], Wipawan Buathong[1,*]

[1] Department of Digital Technology, Faculty of Science and Technology, Phuket Rajabhat University, Phuket 83000, Thailand

* Corresponding author: Wipawan Buathong, w.buathong@pkru.ac.th

Received:
21 January 2024
Revised:
7 April 2024
Accepted:
13 May 2024

Keywords:

Association rule mining, Cybersecurity, Data mining, Phishing awareness, Risk perception

Abstract: This study explores the phishing risk awareness levels among academic staff at Thai universities, employing association rule mining (ARM) to identify critical factors influencing high and low levels of awareness. Targeting a diverse group of 400 academic staff members, the research utilized a structured questionnaire comprising demographic information, direct and indirect experiences with phishing, and perceptions of phishing. In association rules, a lift value of 1 indicates independence between X and Y, while values greater than 1 or less than 1 indicate positive or negative correlation, respectively. The findings revealed several critical findings: despite being able to define phishing, many individuals do not perceive it as a significant threat; moderate internet skills are not necessarily indicative of high phishing awareness; and direct experiences with phishing do not always correlate with an increased awareness of its potential impact. These results highlight a disconnect between knowledge and perceived risk and suggest that existing internet skills and experiences are insufficient for cultivating a robust understanding of phishing risks. The study underscores the necessity for targeted educational interventions specifically designed to address the varied needs of university staff, enhancing their ability to recognize and respond to cybersecurity threats effectively.

1. Introduction

Phishing, a prevalent form of social engineering attack, represents a significant threat to the security of information systems, especially within academic environments (Alabdan, 2020). University academic staff, who have access to sensitive data and play crucial roles in governance, are particularly vulnerable to such attacks. Therefore, assessing their awareness of phishing risks is vital for bolstering institutional cybersecurity measures. Association rule mining, a key data mining technique, is instrumental in extracting meaningful insights from large datasets by revealing interesting relationships between variables (Fister *et al*., 2023). In this study, association rule mining is employed to analyze patterns and correlations concerning the phishing risk awareness levels among university academic staff. This research aims to identify the factors that influence 'phishing awareness = high' and 'phishing awareness = low' among academic staff in Thai universities. This involves investigating how various attributes, determined through a questionnaire, contribute to a high or low level of awareness regarding phishing risks. The selection of university academic staff as the focus of this study is motivated by their critical role in handling sensitive information and their potential impact on shaping cybersecurity policies and practices within academic institutions.

This article explores an in-depth study conducted in Thailand, engaging 400 participants from the academic staff of Thai universities. Carried out in 2020, this research period was notable for a significant rise in cyber threats fueled by increased digital communications. The importance of this study is twofold: its contribution to cybersecurity literature and its focus on a critical group within the academic sector – university staff. Understanding the phishing risk awareness among these individuals is crucial for developing effective strategies to enhance their recognition and response capabilities to cyber threats. Central to this research are three key research questions, all related to the application of association rule mining:

1) What association rules can be identified that indicate high phishing risk awareness among academic staff in Thai universities?

2) What patterns and correlations, as revealed by association rule mining, are associated with low levels of phishing risk awareness in this demographic?

3) How do the identified association rules for high and low phishing awareness inform the development of targeted cybersecurity training and preventive measures within academic institutions?

2. Related Work

Phishing is rapidly increasing as fraud cleverly blends social engineering tactics (Alabdan, 2020). This attack typically involves using deceptive emails or websites designed to dupe unsuspecting victims into divulging personal or sensitive information, leading

to identity theft (Aleroud & Zhou, 2017). Such tactics often escalate to other forms of information security threats. For instance, cybercriminals might use stolen details like account names, passwords, or credit card numbers to create fraudulent credit accounts, committing cybercrimes. Additionally, phishing can involve enticing victims to click on malicious links, resulting in the covert installation of malware on their devices (Parsons *et al*., 2019). The effectiveness of these phishing attacks is often enhanced through sophisticated strategies, including the use of espionage software that aids in gathering critical personal information.

2.1 Phishing Threats and Risk Awareness

Risk perception is central to understanding phishing awareness among academic staff in Thai universities. The APA Dictionary of Psychology (n.d.) defines perception as the awareness of objects, relationships, and events via the senses. In contrast, awareness is the conscious recognition of these elements. Risk involves the likelihood of adverse events and potential loss or harm. Risk perception, or perceived risk, refers to an individual's subjective judgment of the risk associated with a threat (Nam, 2019). These perceptions can be influenced by demographic factors and the sense of control; for example, drivers might perceive lower risk while driving due to their sense of control, as opposed to the high perceived risks of uncontrollable events like earthquakes. Risk perceptions can also be skewed, often underestimated, or exaggerated based on familiarity or lack thereof and are shaped by an individual's knowledge and emotional response to risks (Chavas, 2004).

This research on phishing risk awareness in academic environments parallels these concepts. Phishing, an often uncontrollable and unpredictable threat, is likely to be perceived as a high risk by university staff. However, as with other risks, the perception of phishing threats can be skewed by a lack of awareness of actual statistical data and a tendency to underestimate familiar risks while overestimating unfamiliar ones (Abroshan *et al*., 2021). This ties back to the common assumption in risk perception research: people's knowledge and understanding significantly influence how they perceive risks. Thus, our study investigates the cognitive and emotional dimensions of phishing risk perception among academic staff, exploring how their knowledge, experiences, and feelings toward phishing shape their awareness and preparedness.

With its wealth of personal and institutional data, the academic sector has become a prime target for such attacks (Alharbi & Tassaddiq, 2021; Broadhurst *et al*., 2018; Kenneth *et al*., 2023). Studies have shown that the success of phishing attempts largely depends on the awareness and preparedness of the potential victims, particularly in environments like universities where information exchange is frequent and varied. Several studies have focused on the

general awareness of phishing threats among university staff and students. For instance, Kenneth *et al*. (2023) investigated phishing, a cyber-attack leveraging social engineering to extract sensitive information or prompt clicks on harmful links, often via emails or texts. They utilize an email phishing technique, asking respondents to change their email account passwords, to assess college students' awareness of phishing attacks. The findings reveal that a notable minority remains vulnerable to these attacks, emphasizing the need for awareness campaigns and education to mitigate social engineering risks through phishing.

Broadhurst *et al*. (2018) evaluated the susceptibility of 138 university students to cybercrime through simulated phishing emails, categorizing them into 'Hunter' and 'Passive' groups. Despite varying scam types, factors like cybercrime awareness and IT competence surprisingly did not correlate with susceptibility. The study found tailored emails more effective and international and first-year students more vulnerable. The findings, supported by Generalized Linear Model analysis, emphasize the influence of student status and study year on scam vulnerability, offering directions for future research.

In addition, Alharbi & Tassaddiq (2021) examined cybersecurity awareness among undergraduate students at Majmaah University amid rising technological advancements and associated cyber threats. Utilizing a scientific questionnaire and statistical tests like ANOVA, KMO, and Bartlett's test, they assessed students' understanding of cybercrime and protective measures. The study focused on phishing, computer viruses, and other internet threats. The findings highlighted the need for enhanced education and awareness programs for students to prevent data breaches and digital misconduct.

2.2 Association Rule Mining in Cybersecurity

Association Rule Mining (ARM) has emerged as a powerful tool in cybersecurity, particularly in identifying patterns and predicting potential threats (Tripathi, Nigam, & Edla, 2017). This data mining technique uncovers interesting associations and correlation relationships among large sets of data items (Gu, 2023; Silva *et al*., 2019). In cybersecurity,

A tibble: 400 × 17

Gender <chr>	Age <chr>	Faculty <chr>	Working Period <chr>	Hearing of Phishing <lgl>	Explaining Phishing <lgl>	Defining Phishing <lgl>
M	41 to 45	Science and Technology	6 to 10 years	TRUE	FALSE	TRUE
M	Above 50	Science and Technology	1 to 5 years	TRUE	FALSE	TRUE
F	41 to 45	Science and Technology	6 to 10 years	FALSE	TRUE	FALSE
M	41 to 45	Science and Technology	6 to 10 years	FALSE	TRUE	FALSE
F	46 to 50	Science and Technology	6 to 10 years	FALSE	TRUE	FALSE
M	36 to 40	Science and Technology	1 to 5 years	FALSE	FALSE	TRUE
F	36 to 40	Science and Technology	6 to 10 years	TRUE	FALSE	FALSE
F	31 to 35	Science and Technology	1 to 5 years	FALSE	FALSE	TRUE
M	31 to 35	Science and Technology	1 to 5 years	TRUE	FALSE	FALSE
M	36 to 40	Science and Technology	6 to 10 years	TRUE	FALSE	TRUE

1-10 of 400 rows | 1-7 of 17 columns

Previous 1 2 3 4 5 6 ... 40 Next

Figure 1. Sample data structure: 17 columns and 400 records from phishing awareness survey

ARM has been employed to detect unusual patterns that could signify security breaches, including phishing attacks (Jeeva & Rajsingh, 2016). Studies utilizing ARM in cybersecurity have demonstrated its efficacy in discerning patterns that are not immediately obvious. For instance, Lou *et al*. (2020) applied ARM to detect anomalous behavior in network traffic, successfully identifying potential security threats. In addition, ARM is used to detect unusual patterns that may indicate fraudulent activities on websites (Tripathi, Nigam, & Edla, 2017). Furthermore, Dam *et al*. (2022) addressed the challenge of detecting packer programs, crucial for cybersecurity defenses, by employing associative classification (AC) algorithms. This approach helps classify packers without prior knowledge of feature significance. The study explores and adapts various AC algorithms to manage multiple feature types, assessing their effectiveness in evolving scenarios of packers and malware.

3. Materials and Methods

This section provides a comprehensive overview of the approach.

3.1 Research Design

The study employed a quantitative research approach to analyze the phishing risk awareness levels among academic staff at Thai universities. The primary objective was to identify the factors influencing high and low phishing awareness levels using association rule mining. The data underwent essential design and transformation procedures to ensure its suitability for association rules analysis. These processes involved converting numerical values from rating scales into a categorical format, which resulted in a three-level degree classification: low, moderate, and high. This transformation was instrumental in facilitating the application of association rules analysis, particularly in managing Boolean and categorical data types. The research was structured to collect a wide range of data from participants, encompassing both broad demographic information and specific details regarding their experiences with phishing, their attitudes toward it, and their perceptions of phishing-related risks.

3.2 Data Collection

The study was conducted with a focus on academic personnel from Thai universities, involving a total of 400 participants. The dataset, gathered in 2020, encompasses a diverse group of individuals varying in age, gender, academic discipline, and role within the university. This diversity ensures a comprehensive understanding of the phishing risk awareness levels across different segments of the academic staff. The questionnaire used for data collection comprised 17 questions designed to comprehensively assess various aspects of phishing risk awareness among the academic staff.

These questions can be categorized into three groups based on their nature and the type of responses they elicited:

3.2.1 Demographic Information

Gender, age, faculty, and working period are categorical questions that gather basic demographic information about the participants and provide context for the analysis. It is important to note that the questionnaire did not delve into details specific to university departments. This broad approach allowed for a more general understanding of phishing awareness across various academic roles rather than a detailed analysis of department-specific trends.

3.2.2 Phishing Experience and Attitudes

Hearing of phishing, explaining phishing, defining phishing, being affected by phishing, tackling phishing, experiencing phishing, and observing phishing were structured to elicit 'Yes' or 'No' responses. The focus was on the participants' direct and indirect experiences with phishing and their understanding and response capabilities.

3.2.3 Perceived Phishing Risk

Perceived phishing likelihood, perceived phishing impact, and phishing awareness were designed to assess the participants' perceived likelihood of encountering phishing, the perceived impact of such encounters, and their overall awareness of phishing risks. Responses were classified into three degrees: low, medium, and high. Notably, the phishing awareness degree was derived from a phishing awareness test in the questionnaire, which directly measures the participant's knowledge and awareness level. Figure 1 depicts a sample of the data collection comprising 17 columns and 400 records. Nevertheless, as indicated in Figure 1, a discrepancy is observed in the participants' responses. Some participants assert familiarity with the term 'phishing' yet cannot define it. Conversely, when presented with multiple-choice options, they can select the correct definition of phishing from among ten possibilities. On the other hand, a subset of participants claims to have never encountered the term 'phishing', yet paradoxically, they express confidence in their ability to explain it. However, their responses are incorrect when choosing the appropriate definition from ten options. This inconsistency highlights a potential gap in the participants' understanding and recognition of phishing.

3.3 Research Tool and Ethics Approval

The primary tool for data collection was a structured questionnaire designed to assess various aspects of phishing risk awareness. Questions ranged from essential recognition of phishing attempts to more complex scenarios requiring the identification of subtle phishing cues. The questionnaire was rigorously vetted to ensure clarity, relevance, and unbiased data collection. Before its deployment, the research tool underwent a comprehensive ethical evaluation by the Human and Animal Research Ethics Committee. This full board review, completed on May 10, 2019, emphasized the ethical considerations crucial in research involving human participants. The approval,

valid through May 9, 2020, ensured that the study adhered to the highest research ethics standards, particularly regarding participant consent, confidentiality, and data security.

3.4 Data Analysis

ARM was the primary data analysis technique used in this study. ARM is a method in data mining that identifies interesting relationships, or associations, between variables in large datasets. In the analysis context of this research, ARM was employed to discover patterns in responses that could indicate varying levels of phishing risk awareness among the academic staff. The process of ARM involves three key steps: 1) data preparation (cleaning and organizing the collected data to ensure accuracy and compatibility with the mining process); 2) rule generation (utilizing algorithms to identify frequent itemsets and generate association rules from the data); and 3) rule evaluation (assessing the generated rules to identify the most significant and relevant ones, based on measures like support, confidence, and lift).

The support for each itemset to determine its frequency in the dataset was calculated to identify meaningful association rules. The formula used for calculating support is Support(X)= Number of transactions containing X/Total number of transactions. The support of an itemset X in the transaction dataset T is defined as the proportion of transactions in the dataset that contains the itemset. We then assessed the confidence of each rule to understand the conditional probability of occurrence. The confidence formula applied was Confidence($X \Rightarrow Y$) = Support($X \cup Y$)/Support(X). Confidence of a rule $X \Rightarrow Y$ is defined as the likelihood of finding the itemset Y in transactions because these transactions also contain the itemset X.

Finally, the lift of each rule was computed to measure its effectiveness compared to the baseline probability. The lift is calculated using the formula Lift($X \Rightarrow Y$) = Confidence($X \Rightarrow Y$)/Support(Y). Lift is a measure of the performance of a rule. It compares the rule's confidence with the expected confidence, assuming that the itemsets X and Y are independent. A lift value of 1 indicates the independence of variables X and Y, signifying no association between them. A lift greater than 1 suggests a positive correlation, indicating that the occurrence of X will likely increase the likelihood of Y's occurrence. Conversely, a lift value less than 1 denotes a negative correlation, implying that the presence of X is likely to decrease the likelihood of Y's occurrence. The data analysis was conducted using RStudio, employing essential packages like ggplot2 for visualization, tidyverse for data manipulation, arules for association rule mining, and arulesViz for visualizing the association rules. These tools were instrumental in processing the varied nature of the responses, from categorical and boolean data to the more complex classifications of phishing risk perception.

4. Experimental Results

The application of association rule mining to the dataset revealed several significant patterns and associations related to phishing risk awareness among the academic staff. These findings provide insights into the levels of awareness and the factors influencing them. In applying association rule mining to identify the factors influencing 'phishing awareness = high' and 'phishing awareness = low', a critical adjustment was made in the analysis. For each set of generated association rules targeting either high or low phishing awareness, the rule about the target awareness level itself was excluded. This means that each category's total count of relevant association rules was reduced by one. This adjustment was critical for ensuring that the analysis accurately reflected the actual influences of phishing awareness levels, excluding the target condition itself. The association rule mining procedure was implemented using distinct parameters: a support threshold of 0.033 and a confidence level of 0.85 for high phishing awareness and a support threshold of 0.030 with the same confidence level for low phishing awareness. The rationale for setting higher support for the former was the greater volume of sampling data available for this group. These parameters were selected to guarantee that the derived rules were statistically significant and pertinent to the research objectives.

The analysis was further explored through visualizations created using ggplot 2. Graphs and charts illustrated the distribution of awareness levels and the correlations found in the ARM analysis. The comparative analysis with existing literature indicated that these patterns are consistent with global trends in phishing risk awareness. They also highlight unique aspects specific to the academic environment in Thailand.

4.1 Association Rules for High Phishing Awareness

Table 1 represents a collection of pivotal association rules extracted through data analysis to explore determinants contributing to high awareness of phishing threats. Each rule within the table is articulated through various attributes and quantified by three metrics: support, confidence, and lift. These rules encapsulate a diverse array of demographic and behavioral factors. For instance, the rule combining a low perceived frequency of observing phishing and a high perceived impact of phishing illustrates a robust correlation with elevated phishing awareness, evidenced by a support of 0.035, confidence of 0.93, and a lift of 2.4. Additionally, the data reveal that males possessing moderate perceived internet skills coupled with a high perceived impact of phishing are more likely to demonstrate high awareness of phishing risks, as indicated by a support of 0.045, confidence of 0.86, and lift of 2.2. Moreover, faculty members in the fields of Science and Technology, who have high perceived internet skills and phishing monitoring abilities, along with a high perceived

impact of phishing, show a strong association with high phishing awareness, reflected by a support of 0.040, confidence of 0.89, and a lift of 2.3. These findings underscore the multifaceted influences on phishing awareness and highlight the importance of considering a range of factors when designing interventions to enhance cybersecurity awareness. Table 1 illustrates the nine critical association rules contributing to high phishing awareness levels.

Additional rules presented in the analysis delve into the impacts of age, gender, working period, and direct experiences with phishing, such as hearing about or being personally affected by phishing incidents. Each rule delineated in the study illustrates the intricate factors influencing phishing awareness, emphasizing cybersecurity education's complex and layered aspects and the implications for policy within academic environments. These insights are crucial for developing targeted interventions and tailored training programs to enhance phishing awareness among university staff. Furthermore, these strategies could be adapted to benefit other vulnerable groups within the academic community, thereby broadening the scope of impact and reinforcing overall cybersecurity resilience. Figure 2 displays nine association rules specifically formulated to demonstrate their correlations with high phishing awareness.

Table 1. Nine principal association rules contributing to high phishing awareness

Association Rules	Supp	Conf	Lift
Explaining Phishing, Perceived Phishing Observation Frequency=Low, Perceived Phishing Impact=High	0.035	0.93	2.4
Gender=M, Perceived Internet Skills=Moderate, Perceived Phishing Impact=High	0.045	0.86	2.2
Faculty=Science and Technology, Perceived Internet Skills=High, Perceived Phishing Monitoring Ability=High, Perceived Phishing Impact=High	0.040	0.89	2.3
Age=36 to 40, Working Period=6 to 10 years, Hearing of Phishing, Perceived Phishing Monitoring Ability=High	0.035	0.88	2.3
Working Period=6 to 10 years, Hearing of Phishing, Perceived Phishing Monitoring Ability=High, Perceived Phishing Impact=High	0.040	0.89	2.3
Gender=M, Working Period=6 to 10 years, Hearing of Phishing, Perceived Phishing Impact=High	0.035	0.88	2.3
Working Period=6 to 10 years, Hearing of Phishing, Tackling Phishing, Perceived Phishing Impact=High	0.035	0.93	2.3
Gender=M, Affected by Phishing, Perceived Internet Skills=Moderate, Perceived Phishing Impact=High	0.035	0.93	2.4
Gender=M, Observing Phishing, Perceived Internet Skills=Moderate, Perceived Phishing Impact=High	0.040	0.89	2.4

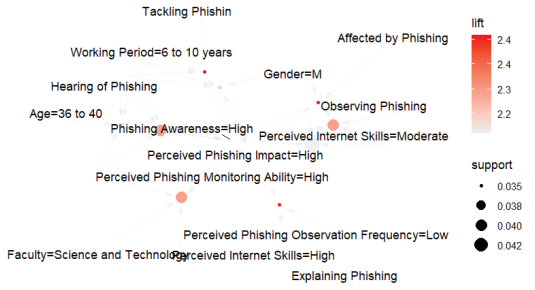

Figure 2. Nine Association Rules Linked to High Phishing Awareness

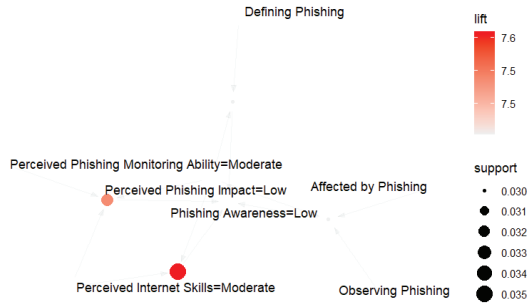

Figure 3. Four Association Rules Correlated with Low Phishing Awareness

4.2 Association Rules for Low Phishing Awareness

Table 2 presents a series of association rules to pinpoint determinants associated with low phishing awareness, where lift values significantly exceeding 1 indicate a robust association. The results from the specific rules elucidated in this table reveal intriguing insights. For instance, individuals may be able to define phishing conceptually yet not regard it as a considerable threat, as indicated by a support of 0.030, a confidence of 0.86, and a lift of 7.5. Furthermore, possessing moderate perceived internet skills does not necessarily align with heightened awareness or concern regarding phishing risks, supported by a lift of 7.6. Additionally, the association between moderate perceived phishing monitoring ability and moderate internet skills further substantiates the correlation with low awareness of phishing's impact, demonstrated by a support of 0.032 and a confidence of 0.87. Notably, individuals who have been affected by phishing and who observe phishing activities may still perceive the impact of phishing as low, suggesting a potential disconnection between personal experiences and risk awareness, with a support of 0.030, a confidence of 0.86, and a lift of 7.5. These findings highlight phishing awareness's complexity and the need for targeted educational approaches to bridge the gap between knowledge and perceived threat levels. Table 2 represents the four essential association rules significantly influencing low phishing awareness levels.

Table 2. Four crucial association rules contributing to low phishing awareness levels

Association Rules	Supp	Conf	Lift
Defining Phishing, Perceived Phishing Impact=Low	0.030	0.86	7.5
Perceived Internet Skills=Moderate, Perceived Phishing Impact=Low	0.035	0.88	7.6
Perceived Internet Skills=Moderate, Perceived Phishing Monitoring Ability=Moderate, Perceived Phishing Impact=Low	0.032	0.87	7.5
Affected by Phishing, Observing Phishing, Perceived Phishing Impact=Low	0.030	0.86	7.5

These association rules highlight critical areas of concern in phishing awareness among university staff. Notably, these findings emphasize the complexity of phishing awareness and suggest that mere exposure to or understanding of phishing does not necessarily translate into a heightened perception of its risks. This table highlights the need for more in-depth and perhaps personalized educational strategies to enhance phishing awareness effectively, particularly among those who might underestimate its potential impact despite having some direct or indirect exposure. Figure 3 presents four association rules specifically established to reveal their correlations with low phishing awareness.

5. Discussion

Phishing attacks often target a broad range of individuals, regardless of their specific field of expertise. Awareness and vulnerability to these threats might be similarly universal across different faculties. In this case, if awareness training and cybersecurity policies are uniformly implemented across the university, irrespective of faculty, this could lead to a leveling effect where faculty affiliation becomes less relevant in determining awareness levels. This finding has important implications for developing phishing awareness programs in academic institutions. It suggests that efforts to enhance phishing awareness could be designed and implemented at the university-wide level rather than tailoring them to specific faculties. Several researchers also supported this view (Alabdan, 2020; Hillman, Harel, & Toch, 2023) and have argued that phishing tests and training can help immunize employees against phishing attacks and reduce undesirable online behaviors. This assertion is substantiated by Kenneth *et al*. (2023), who note that numerous respondents susceptible to social engineering through phishing emails advocate for implementing indoctrination and awareness campaigns targeting students.

5.1 Study Limitations and Future Research

The study focused exclusively on academic staff in Thai universities, which may limit the generalizability of the findings to other demographic groups or geographical locations. Different cultural, educational, or technological environments might yield varying results. In addition, the reliance on self-reported data in the questionnaire raises the possibility of response biases. Participants might have overestimated or underestimated their awareness of phishing risks, impacting the accuracy of the findings.

In addition, the data was collected in a single year (2020), which does not account for potential changes in phishing tactics or awareness over time. Cyber threats evolve rapidly, as do awareness levels and behaviors in response to them. Besides, some questionnaire items, particularly those with boolean (yes/no) responses, may oversimplify the complex nature of phishing awareness. Nuanced understandings or partial awareness might not have been adequately captured.

5.2 Recommendations for Future Research

Future studies could expand the demographic scope to include a broader range of educational institutions, both within and outside Thailand, to test the applicability of these findings in different contexts. Besides, conducting longitudinal studies over several years would provide insights into how phishing awareness evolves and responds to changing cyber threat landscapes and awareness initiatives. Moreover, incorporating qualitative research methods, such as interviews or focus groups, could provide deeper insights into the reasons behind certain awareness levels and the effectiveness of different training approaches. Primarily, investigating the effectiveness of different phishing awareness and training programs would be valuable. This could involve assessing changes in awareness and behavior before and after such interventions.

6. Conclusion

The analysis conducted in this study reveals a nuanced understanding of the factors influencing phishing awareness among Thai university staff. Through the application of association rule mining, several key variables have been identified that significantly affect the levels of phishing awareness. These include demographic characteristics such as age and gender, behavioral factors like perceived internet skills, and direct experiences with phishing activities. The findings indicate that individuals who can conceptually define phishing may not necessarily perceive it as a significant threat. This suggests a disconnect between theoretical knowledge and perceived risk. Moreover, it was found that individuals with moderate perceived internet skills do not consistently demonstrate a heightened awareness or concern about phishing threats, highlighting a gap in the effectiveness of general internet skill levels to foster an understanding of cybersecurity risks.

Additionally, the study uncovered that direct experiences with phishing, such as being affected by or observing phishing attempts, do not always correlate with a high awareness of the impact of such threats. This counterintuitive result points to the need for more comprehensive education and training that increases awareness and enhances the understanding of the consequences of phishing. These findings have significant implications for cybersecurity policy and education within academic settings. They underscore the necessity for targeted educational interventions tailored to the specific needs and experiences of different groups within the university staff. By addressing the multifaceted nature of phishing awareness through customized training programs, institutions can better prepare their personnel to recognize and respond to phishing threats effectively, thereby enhancing the organization's overall cybersecurity posture.

In conclusion, the study demonstrates the value of ARM in understanding complex

behavioral patterns in cybersecurity contexts. The insights gained are crucial for enhancing phishing awareness among academic staff and as a foundational step toward strengthening educational institutions' cybersecurity posture. As cyber threats evolve, continuous research and adaptive strategies will be key in safeguarding sensitive information and maintaining the integrity of academic environments.

CRediT Authorship Contribution Statement

Pita Jarupunphol: Conceptualization, Methodology, Validation, Resources, Software, Formal analysis, Visualization, Writing – original draft, review & editing.

Wipawan Buathong: Model Validation, Resources, Formal analysis, Visualization.

Declaration of Interests

The authors declare that they have no competing financial interests or personal relationships that could have influenced the work reported in this paper.

References

Abroshan, H., Devos, J., Poels, G., & Laermans, E. (2021). Phishing Happens Beyond Technology: The effects of human behaviors and demographics on each step of a phishing process. *IEEE Access, 9*, 44928–44949. https://doi.org/10.1109/access.2021.3066383

Alabdan, R. (2020). Phishing attacks survey: Types, vectors, and technical approaches. *Future Internet, 12*(10), 168. https://doi.org/10.3390/fi12100168

Aleroud, A., & Zhou, L. (2017). Phishing environments, techniques, and countermeasures: A survey. *Computers & Security, 68*, 160–196. https://doi.org/10.1016/j.cose.2017.04.006

Alharbi, T., & Tassaddiq, A. (2021). Assessment of cybersecurity awareness among students of Majmaah University. *Big Data and Cognitive Computing, 5*(2), 23. https://doi.org/10.3390/bdcc5020023

APA Dictionary of Psychology. (n.d.). *APA Dictionary of Psychology*. Retrieved December 17, 2023, from https://dictionary.apa.org/

Broadhurst, R., Skinner, K., Sifniotis, N., & Matamoros-Macias, B. (2018). Cybercrime risks in a university student community. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3176319

Chavas, J.-P. (2004). The measurement of risk. *Risk Analysis in Theory and Practice*, 5–19. https://doi.org/10.1016/b978-012170621-0.50001-8

Dam, K. H. T., Given-Wilson, T., Legay, A., & Veroneze, R. (2022). Packer classification based on association rule mining. *Applied Soft Computing, 127*, 109373. https://doi.org/10.1016/j.asoc.2022.109373

Fister, I., Fister, I., Fister, D., Podgorelec, V., & Salcedo-Sanz, S. (2023). A comprehensive review of visualization methods for association rule mining: Taxonomy, challenges, open problems and future ideas. *Expert Systems with Applications, 233*, 120901. https://doi.org/10.1016/j.eswa.2023.120901

Gu, Y. (2023). Exploring the application of teaching evaluation models incorporating association rules and weighted naive Bayesian algorithms. *Intelligent Systems with Applications, 20*, 200297. https://doi.org/10.1016/j.iswa.2023.200297

Hillman, D., Harel, Y., & Toch, E. (2023). Evaluating organizational phishing awareness training on an enterprise scale. *Computers & Security, 132*, 103364. https://doi.org/10.1016/j.cose.2023.103364

Jeeva, S. C., & Rajsingh, E. B. (2016). Intelligent phishing url detection using association rule mining. *Human-Centric Computing and Information Sciences, 6*(1). https://doi.org/10.1186/s13673-016-0064-3

Kenneth, A., Hayashi, B. B., Lionardi, J., Richie, S., Achmad, S., Junior, F. A., & Nadia. (2023). Phishing attack awareness among college students. *2023 3rd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 344–348. https://doi.org/10.1109/ice3is59323.2023.10335412

Lou, P., Lu, G., Jiang, X., Xiao, Z., Hu, J., & Yan, J. (2020). Cyber intrusion detection through association rule mining on multi-source logs. *Applied Intelligence, 51*(6), 4043–4057. https://doi.org/10.1007/s10489-020-02007-5

Nam, T. (2019). Understanding the gap between perceived threats to and preparedness for cybersecurity. *Technology in Society, 58*, 101122. https://doi.org/10.1016/j.techsoc.2019.03.005

Parsons, K., Butavicius, M., Delfabbro, P., & Lillie, M. (2019). Predicting susceptibility to social influence in phishing emails. *International Journal of Human-Computer Studies, 128*, 17–26. https://doi.org/10.1016/j.ijhcs.2019.02.007

Silva, J., Varela, N., Borrero López, L. A., & Rojas Millán, R. H. (2019). Association rules extraction for customer segmentation in the SMEs sector using the apriori algorithm. *Procedia Computer Science, 151*, 1207–1212. https://doi.org/10.1016/j.procs.2019.04.173

Tripathi, D., Nigam, B., & Edla, D. R. (2017). A novel web fraud detection technique using association rule mining. *Procedia Computer Science, 115*, 274–281. https://doi.org/10.1016/j.procs.2017.09.135