Journal of Computer Science

Original Research Paper

Apriori-Based Analysis of Website Phishing

[1]Rene Clint Gortifacion, [1]Rhoderick Malangsa, [2]Adelfa Diola and [3]Tamar Mejia Junior

[1]Faculty of Computer Science in Information Technology, Southern Leyte State University-Main Campus, Southern Leyte, Philippines
[2]Innovation and Extension Services, Southern Leyte State University-Main Campus, Southern Leyte, Philippines
[3]Faculty of Technical-Vocational Education, Southern Leyte State University-Main Campus, Southern Leyte, Philippines

Article history
Received: 14-08-2024
Revised: 15-11-2024
Accepted: 06-01-2025

Corresponding Author:
Rene Clint Gortifacion
Faculty of Computer Science in Information Technology,
Southern Leyte State University-Main Campus,
Southern Leyte, Philippines
Email: rgortifacion@southernleytestateu.edu.ph

Abstract: Phishing attacks put users' sensitive information at serious risk and are a rising concern in cybersecurity. In order to minimize the possible damage brought on by these assaults, it is essential to detect and categorize phishing websites correctly. In this study, we provide an Apriori-based analysis method for identifying and categorizing website phishing. The Apriori algorithm, which is frequently used in association rule mining, provides a distinctive viewpoint for examining the traits and patterns of phishing websites. This study aims to find significant associations that can help distinguish between legal and phishing websites by using the Apriori algorithm to a dataset of website attributes and related phishing labels. An extensive collection of website labels and attributes, including URL structure, HTML content analysis and other behavioral indicators from UCI, was gathered for the study. We compared the effectiveness of the Apriori-based approach to other phishing detection techniques now in use, such as other machine learning algorithms. In order to create the best rules for this study, the researchers chose to alter the 11,000 datasets run on Weka Software using the Apriori Algorithm. Further, the researchers developed the ten best rules for association on how the Apriori algorithm may be utilized to improve phishing attack detection. This study could improve web security protocols and help prevent phishing attempts, protecting user data and lessening the financial toll of cybercrime.

Keywords: Apriori Algorithm, Cybersecurity, Cybercrime, Weka

Introduction

Phishing is an attack in which the attacker exploits social engineering techniques to steal identity. It traditionally involves sending forged emails, mimicking an online bank, auction, or payment site and guiding users to a bogus web page carefully designed to look like the login to the genuine site. Phishing aims to collect sensitive and personal information such as usernames, passwords, credit card numbers and even money by impersonating a legitimate entity in cyberspace (Aleroud and Zhou, 2017).

Kaspersky Security Network reported that throughout 2022, cybercriminals increasingly turned to phishing. The company's anti-phishing system blocked 507,851,735 attempts to access fraudulent content in 2022, twice as many attacks thwarted in 2021.

Furthermore, Americans experienced a loss of $10.3 billion due to various internet scams in the previous year, as indicated by an FBI report released on March 14, 2023 (Yadav *et al*., 2013).

According to the FBI's annual report, this loss represented the highest amount in five years. The bureau's Internet Crime Complaint Center (IC3) received more than 2,000 daily complaints.

The bureau reported that the most commonly reported crimes were phishing attacks, with 300,497 victims claiming over $52 million in losses in 2022. Phishing, which is defined as "the use of unsolicited emails, text messages and phone calls supposedly from a reputable company requesting personal, financial and login information," tends to be effective because phishing emails often mimic those from known contacts, leading victims to click on insecure links.

Additionally, phishing was the most prevalent type of infection among organizations in Asia in 2021, accounting for 43% of attacks on the continent. It shares this statistic with vulnerability exploitation, while brute force attacks (7%) and the use of stolen credentials (7%) are lower in frequency (https://aag-it.com/the-latest-phishing-statistics/).

Data mining has emerged as a distinct field at the intersection of various disciplines, largely propelled by the growth of extensive databases. The primary motivation for data mining is that these large datasets contain valuable information for their owners (Aburrous *et al*., 2010). However, this information is often concealed within a vast sea of irrelevant data and needs to be revealed. The process involves seeking surprising, novel, or unexpected insights to extract meaningful knowledge. While this field is closely related to exploratory data analysis, the challenges associated with the size of the databases, coupled with concepts and tools borrowed from other domains, suggest that data mining encompasses more than just exploratory data analysis.

The Apriori algorithm is a widely used method in data mining for identifying frequent item sets from transaction datasets and generating association rules (Inokuchi *et al*., 2000). Rules are the knowledge discovered from the database. Identifying frequent item sets (sets of items that appear with a frequency meeting or exceeding a user-defined minimum support) can be challenging due to the combinatorial explosion. After identifying frequent item sets, generating association rules with confidence equal to or greater than a user-specified minimum confidence becomes a straightforward process (Dongre *et al*., 2014).

The Apriori algorithm has been utilized extensively in association rule mining for various purposes, but its ability to associate phishing assaults is still largely untapped (Shayegan Fard and Namin, 2020). Phishing remains a significant challenge due to the constantly evolving characteristics of web pages. It is essential to refine the network architecture to address these changes effectively continuously (Mohammad *et al*., 2014). This particular algorithm was selected for its demonstrated efficacy in uncovering relationships among extensive website attributes, which is critical for identifying patterns indicative of phishing attempts (Dongre *et al*., 2019). In contrast to more intricate machine learning algorithms, the rule-based approach of the Apriori algorithm offers transparency, facilitating direct interpretation of the features correlated with phishing activities.

This study aims to educate IT professionals, businesses, web developers and other people who work in the technology industry about phishing, how to recognize a phishing website and how to help develop policies to combat it based on the most effective rules discovered using the Apriority Algorithm.

Related Work

The study by Senturk *et al*. (2017) stated that social engineering has emerged as a severe threat to virtual communities and is an essential means to attack information systems. The services used by today's knowledge workers prepare the base for complicated social engineering attacks. Phishing is a technically generated social engineering attack. It is the type of identity theft that uses social engineering techniques and complex attack vectors to harvest financial information from unsuspecting consumers (Awatshi and Goel, 2021). It is an attack in which phishers utilize deceptive emails and counterfeit websites to deceive individuals into providing personal information. Victims view these emails as legitimate, whereas they are, in fact, the efforts of phishers looking to commit identity theft. Consequently, there is an immediate need for solutions to combat phishing and several approaches have been identified, along with various strategies to reduce the risk of phishing attacks (Khalili and Sami, 2015). This study proposes a method for detecting phishing by employing machine learning and data mining techniques. A success rate of 89% has been achieved against phishing attacks coming from email messages.

Also, Aung *et al.* (2019) conducted a study to detect phishing attacks using Convolutional Neural Networks and we found that several URL-based features such as the number of subdomains and URL length could also be biased since they highly rely on the dataset. In other words, many researchers use Alexa.com for legitimate datasets, in which only index pages of highly ranked websites are provided. However, phishing datasets from PhishTank.com or OpenPhish.com list all the URLs of the phishing web pages in which phishers use free hosting services that are highly ranked in Alexa. Thus, as for the number of subdomains, legitimate sites from Alexa.com will not have any, while phishing sites will. Furthermore, phishers have complete control over URL composition except for the domain name. Features like URL length can be easily manipulated. Therefore, researchers have recently targeted domain name-based features–instead of entire URL to extract domain name characteristics and current page content.

In addition, Alshahrani *et al.* (2022) used Particle Swarm Optimization and Data Mining to detect phishing and these were the results: The model yields a substantial decrease in the false-positive rates of the phishing URL structure based on the features selected by the classification techniques. Almost all the classifiers have given more than 91% results in identifying the URL phishing under this model. This is a considerable result and it provides more than 98% accuracy in identifying the phishing nature of the URL. The model is sufficient to prove the best results, but more enhanced algorithms from data mining can be applied to the existing model in future work. The study identifies only a limited future for feature selection and the features can be improved. The model is not yet.

Materials and Methods

Apriori algorithm is a type of Data Mining wherein sequence of steps are to be followed to find the most

frequent itemset in the given database. This data mining technique follows the join and the prune steps iterative-ly until the most frequent itemset is achieved. A mini-mum support threshold is given in the problem or it is assumed by the user.

Identify question or goal: In this phase, the researchers determined that the main question is: What is a Phishing Site and what are its characteristics

Collect data samples: The researchers collected desired datasets, in this case, 11,000 datasets from UCI (University of California, Irvine. The UCI Machine Learning Repository is a collection of databases, theories and data generators used by the machine learning community).

Prepare and refine data: The dataset sourced from UCI (University of California, Irvine Machine Learning Repository) was cleaned by normal-izing and converting all attributes into nominal data types. Missing values were handled by substituting default or median values to ensure algorithm stability (Poulis *et al*., 2015). It can be seen in Table (1), the final 29 attributes and description of the dataset used in this study.

Activate apriority algorithm: After cleaning and converting the datasets into nominal data, it is executed through Weka Software with support set at 0.6, 0.8 and 0.9, respectively and Confidence set at 0.9 with three minimum cycles performed. The algorithm determines Confidence for each possible rule, quantifying the chance that one item will be bought in the event another is bought. In addition, it computes lift, a measure of how effective the rule is compared to chance. The choice of support and confidence thresholds impacts the balance between rule coverage and accuracy. Lower support thresholds can uncover more comprehensive patterns but may include noise, whereas higher thresholds focus on strong, consistent patterns at the risk of missing less frequent rules.

Table 1: Matrix of dataset description

Attribute		Description	Data type
1.	Having IP address	Indicates whether the URL contains an IP address in the domain part. Phishing websites sometimes use IP addresses instead of domain names to deceive users	Nominal (Y, N, M)
2.	URL length	Represents the length of the URL. Phishing URLs may have unusually long or short URLs to trick users	Nominal (Y, N, M)
3.	Shortening service	Determines if the URL uses a URL shortening service like Bitly or TinyURL. Such services can hide the actual destination and are commonly used in phishing attacks	Nominal (Y, N, M)
4.	Having the ‘at’ symbol	Check if the URL contains the "@" symbol, which is uncommon in normal URLs but might be used in phishing URLs to deceive users	Nominal (Y, N, M)
5.	Double slash redirecting	Detects if the URL has a double slash ("//") after the scheme (e.g., http://) but before the domain. Phishers may use this technique to redirect users to malicious sites	Nominal (Y, N, M)
6.	Prefix or suffix	Identifies the presence of a prefix or suffix in the domain part of the URL. Phishing URLs may use prefixes or suffixes to mimic legitimate domains	Nominal (Y, N, M)
7.	Having sub domain	Check if the URL contains a subdomain. Phishing URLs may use deceptive subdomains to appear legitimate	Nominal (Y, N, M)
8.	SSL final state	Determines the SSL certificate state of the URL (e.g., secure or not secure). Phishing sites may not have valid SSL certificates	Nominal (Y, N, M)
9.	Domain registration length	Determines the SSL certificate state of the URL (e.g., secure or not secure). Phishing sites may not have valid SSL certificates	Nominal (Y, N, M)
10.	Favicon	Detects the presence of a favicon (a small icon associated with the website). Phishing sites may not have a favicon or use a generic one	Nominal (Y, N, M)
11.	Port	Represents the port number used in the URL. Phishing URLs may use non-standard ports to hide malicious activities	Nominal (Y, N, M)
12.	HTTPS token	Check if "https" is present in the URL. Phishing sites may not use secure connections	Nominal (Y, N, M)
13.	Request URL	Represents the URL used in an HTTP request. It can be used to analyze potentially malicious URLs	Nominal (Y, N, M)
14.	URL of anchor	Refers to the anchor text (hyperlinked text) used in a URL. Useful in detecting phishing links within a web page	Nominal (Y, N, M)
15.	Links in tags	Counts the number of hyperlinks present in the tags of the web page	Nominal (Y, N, M)
16.	SFH (Server From Handler)	Indicates the server form handler used in the URL. Phishing sites may have suspicious form handlers	Nominal (Y, N, M)
17.	Submitting to email	Detects if the URL is submitting data to an email address. Phishing sites may use email submissions for malicious purposes	Nominal (Y, N, M)
18.	Abnormal URL	URLs that have suspicious or abnormal patterns	Nominal (Y, N, M)
19.	Redirect	Whether the URL involves redirection to another page or site, phishers may use redirects to hide their tracks	Nominal (Y, N, M)
20.	On Mouseover	Detects if the URL has a script that runs when the mouse is moved over it. Such scripts might be used in phishing attempts	Nominal (Y, N, M)
21.	Right click event	Check if the URL disables the right-click function on the page. Phishing sites may use this to prevent users from accessing browser features	Nominal (Y, N, M)

22.	Pop-up window	Detects if the URL opens a pop-up window. Phishing sites may use pop-ups to deceive users	Nominal (Y, N, M)
23.	Contains inline frame	Check if the URL contains an inline frame (iframe). Phishers may use iframes to load malicious content	Nominal (Y, N, M)
24.	Age of domain	Represents the age of the domain. Phishing sites often use new domains	Nominal (Y, N, M)
25.	DNS record	Checks if the domain has a valid DNS record. Phishing sites may not have proper DNS configurations	Nominal (Y, N, M)
26.	Web traffic	Refers to the amount of web traffic the URL receives. High web traffic may indicate a popular and legitimate site	Nominal (Y, N, M)
27.	Page rank	Represents the page rank of the URL in search engines. Phishing sites may have low page ranks	Nominal (Y, N, M)
28.	Google index	Indicates if the URL is indexed in Google's search results	Nominal (Y, N, M)
29.	Links pointing to page	Counts the number of external links pointing to the URL. Higher external links may indicate a legitimate website	Nominal (Y, N, M)

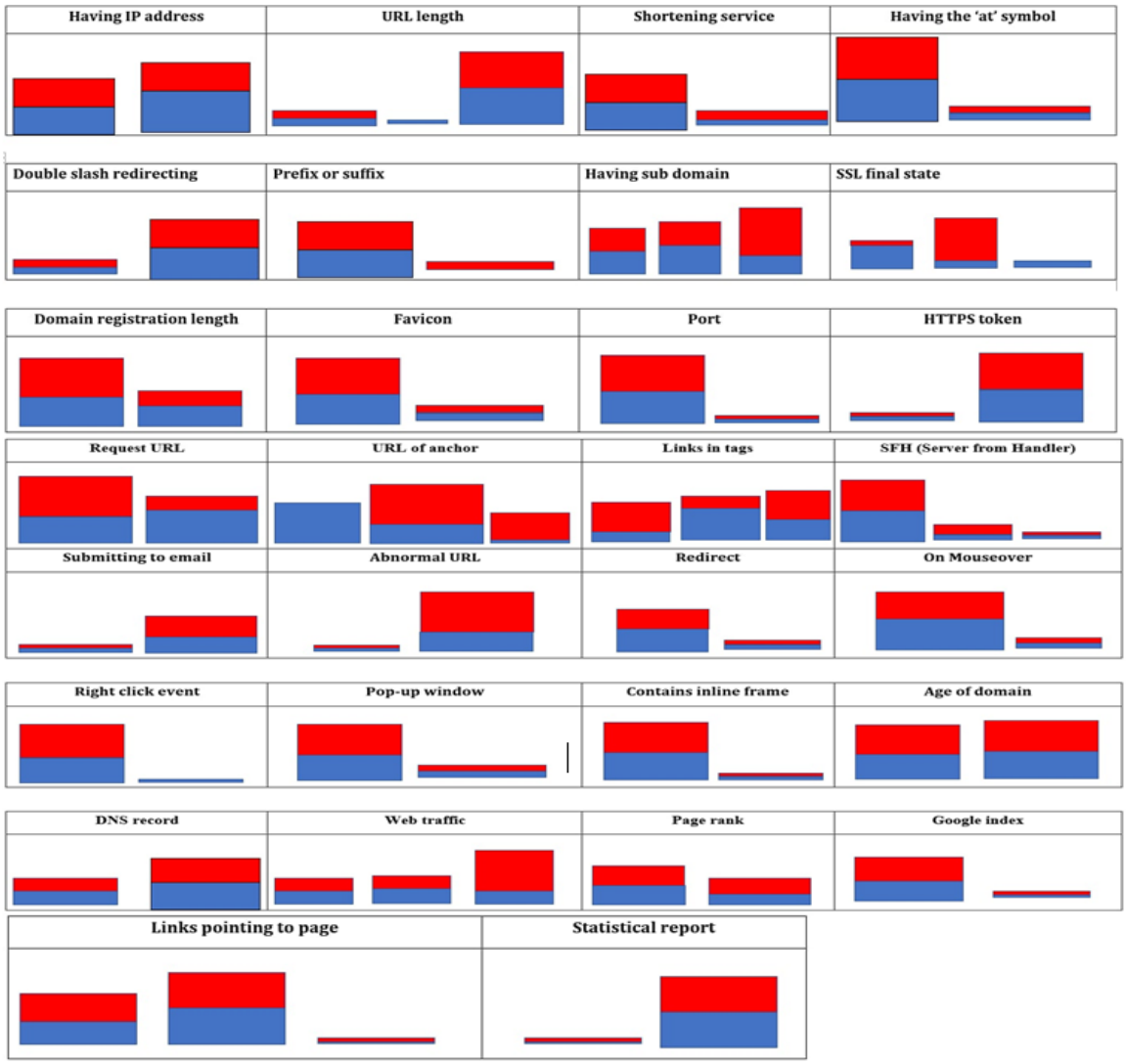

Fig. 1: Visualization of all attributes

There is no overfitting on the datasets, as shown in Fig. (1).

Evaluation: The generated rules were analyzed to understand their practical implications. For instance, rules with high confidence, such as 'A webpage containing both port and an Inline frame will certainly have a Right Click event,' were compared against known phishing cases to validate their reliability. Additionally, the accuracy of the Apriority results was cross-checked against a baseline phishing detection model to assess real-world performance. As seen on Table (2), the confusion matrix is use to evaluate the performance of the model.

Table 2: The confusion matrix

Predicted \ actual	Phishing (1)	Not phishing (0)
Phishing (1)	True positive	False positive
Not phishing (0)	False negative	True negative

1. Having IP Address
2. URL Length
3. Shortening Service
4. Having the 'at' symbol
5. Double slash redirecting
6. Prefix or Suffix
7. Having Sub Domain
8. SSL Final State
9. Domain Registration Length
10. Favicon
11. Port
12. HTTPS Token

1. Request URL
2. URL of Anchor
3. Links in Tags
4. SFH (Server from Handler)
5. Submitting to email
6. Abnormal URL
7. Redirect
8. On Mouseover
9. Right Click Event
10. Pop-up Window
11. Contains Inline Frame
12. Age of Domain
13. DNS Record
14. Web Traffic
15. Page Rank
16. Google Index
17. Links Pointing to Page

This matrix shows that out of 30 websites:

- 10 were correctly identified as phishing
- 15 were correctly identified as not phishing
- 3 were mistakenly labeled as phishing when they were not
- 2 were mistakenly labeled as not phishing when they were

Results and Discussion

After running through the Weka Software with Support set at 0.6, these rules were found: The choice of support and confidence thresholds impacts the balance between rule coverage and accuracy. Lower support thresh-olds can uncover more comprehensive patterns but may include noise, whereas higher thresh-olds focus on strong, consistent patterns at the risk of missing less frequent rules:

1. A webpage containing both a port and an Inline frame will certainly have a right click event
2. A webpage containing an Inline frame will also have a right click event
3. A webpage with a mouse over and an Inline frame will also have a click event
4. If a webpage contains a port, it will certainly also have a right click event
5. A webpage with a mouse over is highly likely to have a right click event
6. A webpage containing both port and right-click events is highly likely to have an Inline frame
7. A webpage containing a port is highly likely to have an Inline frame
8. A webpage containing a port is highly likely to have both a Right Click event and an Inline frame
9. If a webpage has both a mouse over and a right-click event, it is highly likely to have an Inline frame
10. If a web page contains a shortening service, it is highly likely to have a double slash that redirects it to another web page

After running through the weka software with support set at 0.8, these are the rules found: The choice of support and confidence thresholds impacts the balance between rule coverage and accuracy. Lower support thresholds can uncover more comprehensive patterns but may include noise, whereas higher thresholds focus on strong, consistent patterns at the risk of missing less frequent rules:

1. A website with both port and Iframe will have right click
2. A website with an Iframe will likely have right click, too
3. A website with both on mouse over and Iframe will likely have right click
4. A website with a port is highly likely to have right click, too
5. If a website has on mouse over, it will likely have right click too
6. A website with both port and right click will likely have an Iframe
7. A website with a port is very likely to have an Iframe, too
8. A website with a port and right click will likely have an Iframe, too
9. A website with both on mouse over and right click will likely have an Iframe, too
10. If a website uses a URL shortening service, it will likely have double slash redirecting

After running through the Weka Software with Support set at 0.9, these are the rules found: The choice of support and confidence thresholds impacts the balance between rule coverage and accuracy. Lower support thresholds can uncover more comprehensive patterns but may include noise, whereas higher thresholds focus on strong, consistent patterns at the risk of missing less frequent rules:

1. If a website has iframe, it is almost certain that it will also have right click
2. If a website has right click, there is a 95% chance that it will also have Iframe

After Summarization of the Rules Found with Their Respective Confidence Set, These Are the Most Consistent Rules Gathered by The Researchers and Was Considered the Best Rules

A webpage containing both a port and an Inline frame will certainly have a Right Click event:

When a webpage contains both the "port" and "Inline frame" attributes, it uses a specific internet channel and displays content from another website. The rule suggests that in such cases, there is a high probability that the webpage will also have the functionality to allow users to right-click on the page

A webpage containing an Inline frame will also have a Right Click event:

When a webpage contains the "Inline frame" attribute, it displays content from another website within itself. The rule suggests that in such cases, the webpage will likely also have the functionality to allow users to right-click on it

If a webpage has both an on mouse over and an Inline frame, it is highly certain that it will also have a Click event:

A webpage with both the "on mouse over" and "Inline frame" attributes has interactive elements where something happens when you move your mouse over them. It is also displaying content from another website within itself:

A webpage containing a port will certainly have a Right Click event:

A webpage containing the "port" attribute means it uses a specific internet channel for communication.
The rule suggests that in such cases, the webpage will likely also have the functionality to allow users to right-click on it

If a webpage has on mouse over, it is highly likely to have a Right Click event:

The "on mouse over" attribute indicates that the webpage will likely have a "Right Click event" feature, where users can right-click on elements of the webpage to access additional options or actions

A webpage containing both port and Right Click event is highly likely to have an Inline frame:

The presence of both "port" and "Right Click event" together strongly indicates that the webpage will likely have an "Inline frame" feature, where it can show content from other websites within the current webpage

A webpage containing a port is highly likely to have an Inline frame:

A webpage containing the "port" attribute uses a specific internet channel for communication
The rule is suggesting that in such cases, there's a high probability that the webpage will also have an "Inline frame," which means it will be able to display content from other websites within itself

A webpage containing a port is highly likely to have both a Right Click event and an Inline frame:

The presence of the "port" attribute strongly indicates that the webpage will likely have both a "Right Click event" and an "Inline frame" feature, making it interactive with the ability to display external content

A webpage with an on mouse over and Right Click event is highly likely to have an Inline frame:

The presence of both "on mouse over" and "Right Click event" together strongly indicates that the webpage will likely have an "Inline frame" feature, making it interactive and capable of displaying external content

These are the formula and the calculation for the metrics.

Accuracy: The proportion of total correct predictions (both true positives and true negatives) out of all predictions:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision (for phishing): The proportion of correctly predicted phishing websites out of all predicted phishing websites:

$$\text{Pr}ecision = \frac{TP}{TP + FP}$$

Recall (for Phishing): The proportion of correctly predicted phishing websites out of all actual phishing websites:

$$\text{Re}call = \frac{TP}{TP + FN}$$

F1 Score: The harmonic mean of precision and recall, providing a balance between the two:

$$F1 = 2 \times \frac{\text{Pr}ecision \times \text{Re}call}{\text{Pr}ecision + \text{Re}call}$$

We utilized the formulas mentioned to achieve the following results:

Table 3: Based on the confusion matrix

Predicted \ actual	Phishing (1)	Not phishing (0)
Phishing (1)	10 (True positive)	3 (False positive)
Not phishing (0)	2 (False negative)	15 (True negative)

$$Accuracy = \frac{10 + 15}{10 + 15 + 3 + 2} = \frac{25}{30} \approx 0.8333 \; or \; 83.33\%$$

$$\text{Pr}ecision = \frac{10}{10 + 3} = \frac{10}{13} \approx 0.7692 \; or \; 76.92\%$$

$$\text{Re}call = \frac{10}{10 + 2} = \frac{10}{12} \approx 0.8333 \; or \; 83.33\%$$

$$F1 = 2 \times \frac{0.7692 \times 0.8333}{0.7692 \times 0.8333} \approx 2 \times = \frac{0.6401}{1.6025} \approx 0.7989 \; or \; 79.89\%$$

As shown in Table (3), these are the results:

- 10 were correctly identified as phishing
- 15 were correctly identified as not phishing
- 3 were mistakenly labeled as phishing when they were not
- 2 were mistakenly labeled as not phishing when they were

Summary of metrics:

Accuracy	:	83.33%
Precision	:	76.92%
Recall	:	83.33%
F1 Score	:	79.89%

Conclusion

The model correctly classified a large majority of the websites, whether phishing or non-phishing. This indicates overall good performance with 83.33% and also, with the aid of Weka Software, the researchers were able to successfully identify the best rules based on the analysis of the datasets and these rules could be used to educate and be very helpful to both Internet Security Providers and Web Developers in order to create policies regarding Website Phishing, through which Website Phishing would be mitigated.

Acknowledgment

Our thanks to southern Leyte state university for this opportunity.

Funding Information

The authors should acknowledge the funders of this manuscript and provide all necessary funding information.

Author's Contributions

Rene Clint Gortifacion: Concept of work, presentation and methodology, choice of material.

Rhoderick Malangsa: Concept of work; presentation and methodology; supervision; final proofreader.

Adelfa Diola: Written, reviewed, edited; proofreader.

Tamar Mejia Junior: Written, reviewed, edited; typesetter.

Ethics

Authors declare no conflict of interest. Also, all materials used in this article particularly images, are available online without copyright issues.

References

Aburrous, M., Hossain, M. A., Dahal, K., & Thabtah, F. (2010). Predicting Phishing Websites Using Classification Mining Techniques with Experimental Case Studies. *2010 Seventh International Conference on Information Technology: New Generations*, 176–181. https://doi.org/10.1109/itng.2010.117

Aleroud, A., & Zhou, L. (2017). Phishing Environments, Techniques, and Countermeasures: A Surveya. *Computers and Security*, *68*, 160–196. https://doi.org/10.1016/j.cose.2017.04.006

Alshahrani, S. M., Khan, N. A., Almalki, J., & Shehri, W. A. (2022). URL Phishing Detection Using Particle Swarm Optimization and Data Mining. *Computers, Materials & Continua*, *73*(3), 5625–5640. https://doi.org/10.32604/cmc.2022.030982

Aung, E. S., Zan, C. T., & Yamana, H. (2019). A Survey of URL-Based Phishing Detection. *DEIM Forum*, G2-3.

Awasthi, A., & Goel, N. (2021). Phishing Website Prediction: A Machine Learning Approach. *Progress in Advanced Computing and Intelligent Engineering*, *1299*, 143–152. https://doi.org/10.1007/978-981-33-4299-6_12

Dongre, J., Prajapati, G. L., & Tokekar, S. V. (2014). The Role of Apriori Algorithm for Finding the Association Rules in Data Mining. *2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, 657–660. https://doi.org/10.1109/icicict.2014.6781357

Inokuchi, A., Washio, T., & Motoda, H. (2000). An Apriori-Based Algorithm for Mining Frequent Substructures from Graph Data. *Principles of Data Mining and Knowledge Discovery*, *1910*, 13–23. https://doi.org/10.1007/3-540-45372-5_2

Khalili, A., & Sami, A. (2015). SysDetect: A Systematic Approach to Critical State Determination for Industrial Intrusion Detection Systems using Apriori Algorithm. *Journal of Process Control*, *32*, 154–160. https://doi.org/10.1016/j.jprocont.2015.04.005

Mohammad, R. M., Thabtah, F., & McCluskey, L. (2014). Predicting Phishing Websites Based on Self-Structuring Neural Network. *Neural Computing and Applications*, *25*(2), 443–458. https://doi.org/10.1007/s00521-013-1490-z

Poulis, G., Gkoulalas-Divanis, A., Loukides, G., Skiadopoulos, S., & Tryfonopoulos, C. (2015). SECRETA: A Tool for Anonymizing Relational, Transaction and RT-Datasets. *Medical Data Privacy Handbook*, 83–109. https://doi.org/10.1007/978-3-319-23633-9_5

Senturk, S., Yerli, E., & Sogukpinar, I. (2017). Email Phishing Detection and Prevention by Using Data Mining Techniques. *2017 International Conference on Computer Science and Engineering (UBMK)*, 707–712. https://doi.org/10.1109/ubmk.2017.8093510

Shayegan Fard, M. J., & Namin, P. A. (2020). Review of Apriori based Frequent Itemset Mining Solutions on Big Data. *2020 6th International Conference on Web Research (ICWR)*, 157–164. https://doi.org/10.1109/icwr49608.2020.9122295

Yadav, C., Wang, S., & Kumar, M. (2013). An Approach to Improve Apriori Algorithm Based on Association Rule Mining. *2013 4th International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 1–9. https://doi.org/10.1109/icccnt.2013.6726678