Unmasking Phishing Attempts
A Study on Detection in Spanish Emails

Herrera-Semenets, Vitali; Bustio-Martínez, Lázaro; Pérez-Guadarramas, Yamel; González-Ordiano, Jorge Ángel; van den Berg, Jan

DOI
10.1007/978-3-031-76604-6_1

Publication date
2025

Document Version
Final published version

Published in
Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications

Unmasking Phishing Attempts: A Study on Detection in Spanish Emails

Vitali Herrera-Semenets[1], Lázaro Bustio-Martínez[2](✉), Yamel Pérez-Guadarramas[1], Jorge Ángel González-Ordiano[3], and Jan van den Berg[4]

[1] Advanced Technologies Application Center (CENATAV), La Habana, Cuba
{vherrera,yperez}@cenatav.co.cu

[2] Departamento de Estudios en Ingeniería la Innovaciniversidad Iberoamericana
Ciudad de México, Mexico City, Mexico
lazaro.bustio@ibero.mx

[3] Instituto de Investigación Aplicada y Tecnología Universidad Iberoamericana
Ciudad de México, Mexico City, Mexico
jorge.gonzalez@ibero.mx

[4] Intelligent Systems Department, Delft University of Technology,
Delft, The Netherlands
j.vandenberg@tudelft.nl

Abstract. Phishing, a pervasive cybersecurity issue, involves fraudulent attempts to obtain sensitive information and to provoke unintentional money transfers or malware downloads, among others, by disguising as trustworthy entities in electronic communications. This paper presents an innovative approach to phishing detection in Spanish emails using patterns represented as rules. Through a comprehensive, still efficient analysis of emails, we identify interpretable recurring patterns and relevant phrases used in phishing attempts. These phrases and words often aim to persuade victims into revealing personal or financial information. These patterns are translated into a set of rules that are applied to evaluate incoming emails. Additionally, a proof-of-concept is carried out using a phishing data set of Spanish emails created for this study. Our method achieved promising results in identifying phishing attempts, providing an additional layer of security for email users. Moreover, this approach can be adapted to detect phishing in other languages, making it a potentially global solution to this persistent cybersecurity issue.

Keywords: Machine learning · Phishing detection · Rule-based systems · Spanish emails

1 Introduction

Phishing is a pervasive cybersecurity issue that continues to plague individuals and organizations worldwide. It involves fraudulent attempts to obtain sensitive information such as usernames, passwords, credit card details and to provoke

R. J. Barrientos and S. A. Velastin (Eds.): CIARP 2024, LNCS 15369, pp. 1–15, 2025.
https://doi.org/10.1007/978-3-031-76604-6_1

unintentional money transfers or malware downloads by disguising as a trustworthy entity in an electronic communication.

In 2022, Europol reported the arrest of a phishing gang that was behind losses worth several million euros in Belgium and the Netherlands [11]. The modus operandi involved sending emails and text messages containing phishing links to fake banking websites. Thinking they were viewing their own bank accounts through this website, the victims were duped into providing their banking credentials to the suspects. According to Proofpoint's annual "State of Phish" report published in 2023 [20], the 90% of Spanish companies surveyed experienced a successful phishing attack via email. Consequently, a quarter of them recorded financial losses.

The problem is further exacerbated when the phishing attempts are made in languages other than English, such as Spanish, where detection tools are less developed and public databases for phishing detection are scarce. This lack of resources presents a significant challenge in developing effective phishing detection systems for Spanish emails, leaving a large population of internet users vulnerable to these attacks.

Motivated by this gap, this paper introduces an innovative approach to tackle this issue. We present a method for phishing detection in Spanish emails using patterns represented as rules. Through a comprehensive analysis of emails, we identify recurring patterns and key phrases used in phishing attempts. These phrases and words often aim to persuade victims into revealing sensitive information or perform an unintentional action. To extract the key phrases, 212 samples of different phishing attempts, in Spanish emails, were collected. The patterns identified from the data set are then translated into a set of rules that are applied to evaluate incoming emails.

Our method demonstrates high efficacy in identifying phishing attempts. Additionally, the proposed method can provide an additional layer of security for email users by integrating it into existing defensive tools. Moreover, our approach can be adapted to detect phishing attempts in emails written in other languages, making it a potentially global solution to this persistent cybersecurity issue.

The remainder of this paper is structured as follows. First, the background about the motivation underlying this research, phishing detection in Spanish emails, is described in Sect. 2. Second, the proposed strategy is presented in Sect. 3. Third, in Sect. 4, the experimental results using different settings are discussed. Finally, the conclusions and future work are outlined in Sect. 5.

2 Background

A wide variety of proposals for phishing detection based on machine learning techniques are reported in the literature [18]. Figure 1 shows a taxonomy that covers the most addressed approaches.

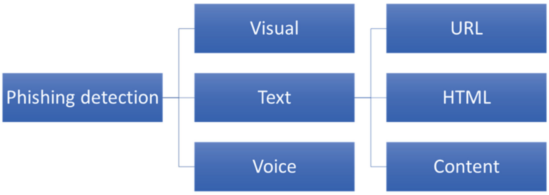


Fig. 1. Taxonomy of the most addressed approaches for phishing detection.

The analysis of visual information can range from the processing of different types of images, including logos or even favicons[1] [6,15]. These approaches are often expensive, in terms of temporal and spatial complexity, since it requires tracing the images to identify visual elements that contribute to extract patterns that describe a phishing attempt. Furthermore, large storage capacities are required and large-scale use is not feasible.

Voice Phishing, also called vishing or phone call scam, is another way to steal people's data and information. The proposals in this area have focused mostly on sentiment analysis, the content of the conversation and verification of the call originating number. To do this, the use of techniques based on natural language processing [5], machine learning [16], and deep learning [17,30], has been more common. Vishing detection is an area in which more promising results are beginning to be seen in recent years. However, there are still several limitations that make people very vulnerable to vishing. For example, difficulty tracking and verifying numbers, emotional manipulation, use of deepfakes, and the ability to detect in real time. The latter is very important, since being faced with a text that may be suspicious, where a person can take time to decide whether to respond or not, is not the same as being in the middle of a phone call, where the attacker permanently insists on obtaining information immediately.

The analysis of the information available in text has been approached from multiple perspectives, the most popular being: URL[2], HTML and content-based. URL and HTML analysis has a wide variety of proposals, mostly focused on detecting phishing websites. Wang et al. [28] uses Internationalized Domain Names (IDN) analysis to detect phishing attempts. To do this, they transform domain names into images and calculate their similarity by Siamese neural networks. This way they can identify whether a domain name is IDN homograph or not. A lightweight data representation for phishing URLs detection in IoT environments was proposed in [7]. Here, the combination of a novel feature selection algorithm and a classifier based on decision trees, led them to achieve an efficacy above 99% in the detection of phishing URLs. There is another large number of reported works, which are based on neural networks [10,22,29].

[1] A favicon, or "favorite icon", is a small 16 × 16 pixel icon used in web browsers to represent a site or web page.

[2] URL is an acronym for Uniform Resource Locator and is a reference to a unique resource on the Internet.

A content-based feature extraction was proposed by Bountakas and Xenakis [4]. The authors argued that the contents of emails contain information that can be employed for phishing email detection. More specifically, the proposed methodology supports that phishing emails follow a particular format, where their contents differ from benign emails. Based on the format of emails as well as the phishing email traits, 4 feature categories have been used:

- Phishing emails are designed to perform particular malicious actions (e.g., deceiving the victim to share personal information or click a malicious attachment). To accomplish these actions, phishing emails employ Body Features that are related to the body structure of emails.
- The text of phishing emails is comprised of strong persuasion traits that are represented by Syntactic Features from the emails body and subject fields.
- The emails include a header field that generates Header Features, which provide useful information about the emails type and the number of recipients.
- URL Features obtained from the hyperlinks and the domains of emails to indicate whether the email contains malicious hyperlinks or is generated by peculiar domains.

Content analysis is a widely addressed approach, mainly in electronic messages or emails. In these cases, it is usually common to use methods based on semantic analysis and natural language processing to analyze text and detect inappropriate statements that are indicative of phishing attacks [1,3]. The use of deep neural networks is also noticeable in this area [26]. Hiransha et al. [14] utilized Keras Word Embedding and Convolutional Neural Networks (CNN) to construct their model, which aims to differentiate phishing emails from legitimate ones. The experimental results reported in [21], show that deep learning word embedding, specifically Long Short-Term Memory (LSTM), is appropriate for the email anti-phishing task.

Regardless of the area addressed, the marked tendency to use techniques based on deep learning that lead to good results is notable. Specifically, the CNN together with the LSTM networks are the most used in these scenarios [2,27,31]. However, people often face phishing attempts almost daily and there is not always a mechanism at hand, whether based on deep learning or not, that helps in making decisions in a situation of this type. The interpretability of specific patterns or characteristics that can lead a person to intuit that they are in the presence of a phishing attempt is very limited when deep learning techniques are used.

On the other hand, there are approaches, such as rule-based ones, that are easy to understand and explain. Additionally, rule-based methods do not require large volumes of data for their development, unlike deep learning systems that often need large data sets to train effectively. This aspect is very important when we face phishing attempts in languages other than English, where detection is usually more complex due to the lack of available data sets that allow training models. According to [24], Spanish is the 4th most spoken language in the world and the second with the largest number of native speakers (see Fig. 2). This situation represents a substantial number of individuals who may be exposed to or at risk of becoming victims of phishing, making this group a common target for phishing attempts.

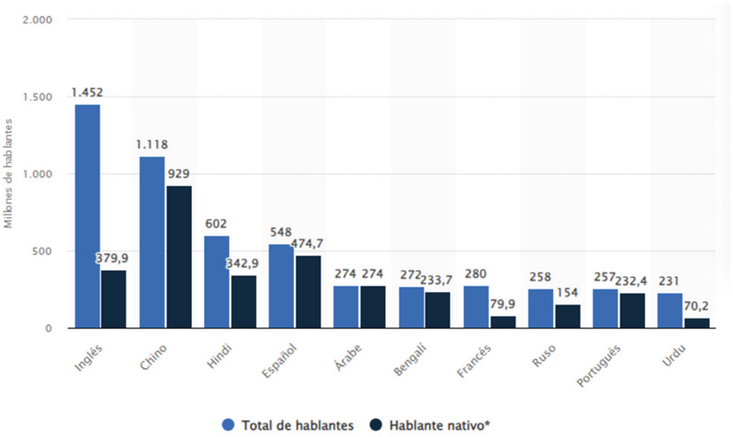


Fig. 2. The most spoken languages worldwide in 2023 [24].

From the previous analysis it is possible to formulate the following reasons that will guide our work:

- **Language coverage:** Although many phishing detection systems are developed and tested primarily in English, phishing is not limited to any particular language. A system that can effectively detect phishing in Spanish could protect a large number of Spanish-speaking users.
- **Cultural adaptability:** Phishing attacks are often adapted to local cultures and contexts to be more convincing. A system that understands the specific rules and patterns of Spanish could be more effective in detecting these culturally adapted attacks.
- **Speed and efficiency:** As mentioned before, rules-based systems can be faster and more efficient than deep learning-based systems. This could allow for faster detection and response to phishing attempts.
- **Interpretability:** A rules-based system can provide a clear explanation of why a message was flagged as phishing. This can be useful for user education and improving trust in the system.

Taking the above into account, we can deduce that it would be very beneficial to have a rule-based mechanism to detect phishing in Spanish. Therefore, in this work a rule-based strategy is proposed to identify, with a high accuracy, the presence of a phishing attempt in Spanish. The purpose of having a rules-based strategy is not only to detect phishing, but also to contribute to user education and awareness in preventing phishing.

3 Proposal

The proposed strategy consists of four fundamental steps (see Fig. 3).

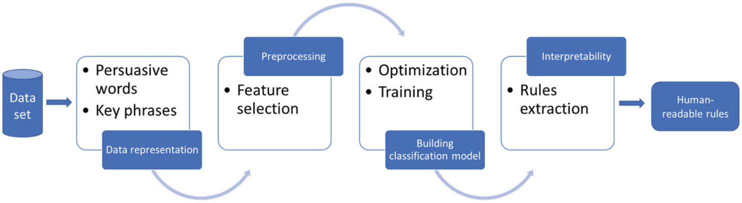


Fig. 3. General scheme of the proposed strategy.

The first step is to represent the data as persuasive words and key phrases. Persuasive words are those used to convince, influence or induce someone to take an action or change their point of view. These words can provoke specific emotions and push potential victims to certain actions. These actions may have various purposes, such as political, marketing and, especially, phishing. These words are powerful tools in the art of copywriting, which is writing texts in such a way that they provoke a direct reaction in the reader. Figure 4 shows some examples of persuasive words grouped into five categories.

Trust	Urgency	Scarcity	Added value	Curiosity
• Easy • Guaranteed • Safe	• Instant • Quick • First	• Only • Limited	• Free • Gift • Bonus	• New • Secrets • Tricks

Fig. 4. Examples of persuasive words.

The key phrases are those commonly used in phishing emails to trick users. These phrases may include, for example, urgent requests for personal information, promises of financial rewards, threats of negative consequences if immediate action is not taken, among others. The key phrases extraction algorithm used in this strategy operates in 5 stages: Text-Preprocessing, Candidate Phrases Extraction, Topics Identification, Topics Ranking Construction and Key-phrases Selection. For more details about this algorithm, we refer to the work described in [19]. In this way, each message is represented as an n-dimensional vector, where n is the total number of persuasive words and key phrases. Each position of the vector is assigned a binary value, indicating the presence (1) or not (0) of the word or phrase. This representation captures the essence of phishing content, which often includes persuasive and relevant language to trick users. A characteristic of this data representation is that a high-dimensional data set is obtained, where not all the features provide useful information for the training process.

Therefore, the second step involves feature selection. For this, an algorithm that has been shown to be feasible for intrusion detection task was used [13]. In essence, the algorithm involves using three measures: Chi-Squared, Information Gain, and ReliefF. Each measure leverages different types of qualitative information to select features. After obtaining a ranking of scores with each measure, the mean score is computed for each one. This mean serves as a threshold for selecting features above that value. Then, by combining the selected feature subsets, the final feature set is obtained.

The third step concerns the building of the classification model. Here three classifiers were evaluated, from which rule-based patterns can be obtained. Such classifiers are RIPPER [9], IREP [12] and DecisionTreeClassifier [23]. Section 4 shows that the latter best suited this context.

Algorithm 1: Building classification model

```
    Input: $\tau$: Threshold, $\rho$: hyper-parameters, $\zeta$: classifier, $D$: preprocessed dataset
    Output: $\zeta_M$: Classification Model
 1  $scores = [Accuracy, Recall, Precision]$
 2  $models = GridSearch(\zeta, \rho, scores, D, cv = 10)$
 3  foreach $model$ in $models$ do
 4      if $model.Accuracy > \tau$ then
 5          selectedModels.Add(model)
 6      end
 7  end
 8  foreach $model$ in $selectedModels$ do
 9      if $model.Recall > (selectedModels.MaxRecall() - selectedModels.StdRecall())$
          then
10          filteredModels.Add(model)
11      end
12  end
13  $bestModel$ = filteredModels.Get_Model_by($Max\_Precision$)
14  $\zeta_M$ = TrainModel($\zeta, D, bestModel.Optimal\_Values(\rho)$)
15  return $\zeta_M$
```

As can be seen in line 2 of Algorithm 1 shown below,the building classification model step involves applying Grid Search to optimize the model hyper-parameters based on three quality measures: Accuracy (see Eq. 1), Recall (see Eq. 2) and Precision (see Eq. 3). The values of the quality measures reported in this process were obtained by also applying a 10-fold cross-validation during the Grid Search process. To select the best combination of hyper-parameters, a strategy was established based on a defined threshold for the Accuracy measure. Those built models, during Grid Search, that report an accuracy above the threshold are selected as candidates (see lines 3–7 of Algorithm 1). Then, those models whose Recall is greater than the maximum reported Recall minus the standard deviation are filtered (see lines 8–12 of Algorithm 1). From the filtered models, the one whose precision is the highest is finally selected (see line 13 of Algorithm 1). Once the model with the best combination of hyper-parameters is selected, such hyper-parameters are used to train the classifier using the preprocessed dataset (see line 14 of Algorithm 1).

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Predictions}} \tag{1}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{2}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{3}$$

The fourth and final step involves rule extraction. In this step, human-readable rules are extracted from the classification model. Although there are models that allow patterns represented as rules to be directly obtained, there are others, such as those based on decision trees, of which it is also possible to represent them in rule structures. For this case, the idea of representing a rule as a path from the root node to a leaf node was followed. This allowed us to obtain a set of rules made up of all possible paths from the root node to each of the leaf nodes of the decision tree. These rules help to understand how the model makes decisions and provide transparency, which is especially important in security applications such as phishing detection.

This proposal offers a systematic and understandable approach to detecting phishing in emails, from the initial representation of the data to the interpretation of the model results.

4 Experimental Results

The non-availability of a data set of phishing emails in Spanish led us to create a first approximation of what could be a reference data set to evaluate approaches aimed at detecting phishing in Spanish. It is important to note that there is a tendency in the literature to assume that a spam email is the same as a phishing email. This assumption is wrong because phishing attacks involve targeted attempts to trick people into revealing sensitive information or taking specific actions, while spam typically encompasses unsolicited mass messages. Combining both types of emails can lead to inaccuracies in detection, as phishing emails often mimic legitimate communications and rely heavily on psychological manipulation to attempt to promote certain emotions in their victims using Principles of Persuasion [8], as opposed to spam. Additionally, the different characteristics and intent of phishing attacks require customized detection methods that go beyond traditional spam filters.

Taking the above into account, 212 phishing emails in Spanish were collected. To balance the data set, 212 legitimate emails were also collected. Finally, the data set was made up of a total of 424 emails. This data set was collected through the joint work of 2 institutions: IBERO and CENATAV. The phishing messages collected primarily include attempts to obtain sensitive personal and financial information, such as account numbers, usernames and passwords. A part of these emails seek to get the victim to access spoofing websites by posing as bank officials, legal authorities, or online site administrators. In other cases, they request the information directly by email. However, a common aspect in these emails is the use of certain persuasive words and relevant phrases that seduce the victim to follow the scammer's game.

By applying the first step of the proposed strategy, 687 key phrases were extracted from the training set. Additionally, 145 persuasive words were used to represent the data. This adds up to a total of 832 words and phrases, of which 13 are repeated, which is why they were removed. Therefore, each email was represented as a vector of 819 dimensions, where each dimension is associated with a phrase or word.

It is important to note that not all features used to represent data are necessarily useful or provide relevant information. Therefore, it is essential to apply a feature selection process. This process allows us to identify and select those features that significantly contribute to our analysis, thereby improving the efficiency and accuracy of our data models. In the next step, the feature selection was carried out. This allowed us to reduce the dimensionality of the data to 25 features.

The next step consists of applying a Grid Search process to obtain the optimal values of the hyper-parameters. This procedure was implemented on three distinct classifiers, enabling the acquisition of rule-based pattern representations: RIPPER, IREP and Decision Tree Classifier. For the latter, the fourth step of the proposed strategy was applied, which allows building rules from the created decision tree.

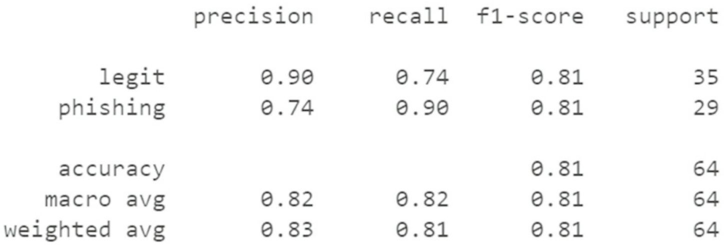
As described in the previous section, a heuristic to select the optimal combination of values for the hyper-parameters was designed. In this sense, for each classifier, the model built that reported the highest accuracy, during the Grid Search process, was selected. In the case of IREP and RIPPER, the efficacy achieved, in terms of accuracy, was 0.65 and 0.66 respectively; somewhat lower than the 0.74 achieved with Decision Tree Classifier. This result led to using Decision Tree Classifier as the model for building rules from the proposed data representation.

In order to evaluate the rules obtained with Decision Tree Classifier, the data set was divided into a training set (85%) and a test set (15%). Once the classifier was trained, from the decision tree built, it was possible to create 20 rules (12 legitimate rules and 8 phishing rules). Below is an example of a phishing rule obtained.

- if (comunidad = 0) and (saludo = 0) and (investigador = 0) and (caso = 0) and (correo = 1) and (cuenta = 1) and (clic = 1) then class: **phishing** (proba: 100.0%) | based on 14 samples

This rule evaluates the presence, or absence, of certain words or phrases in a text, in such a way that they define a pattern associated with a phishing attempt. The rule shown has a probability, or confidence, of 100% based on 14 samples that were covered during training and that matched the phishing class that this rule represents.

The classification model demonstrated a robust performance with an accuracy of 0.81. Notably, it achieved a promising recall of 0.90 for the phishing class, contributing to an average recall of 0.82 (see Fig. 5). Although the data set used is still small to be able to estimate the behavior of the rules obtained in a real-world context, the results are encouraging.



```
              precision    recall  f1-score   support

       legit       0.90      0.74      0.81        35
    phishing       0.74      0.90      0.81        29

    accuracy                           0.81        64
   macro avg       0.82      0.82      0.81        64
weighted avg       0.83      0.81      0.81        64
```

Fig. 5. Achieved results with Decision Tree Classifier.

In Fig. 6 it can be seen that only 3 emails were false negatives and 9 false positives. A false negative occurs when a real phishing attack goes undetected and is allowed to continue. This can lead to serious consequences, such as the theft of personal or financial information. Therefore, it is essential that phishing detection systems are able to correctly identify most, if not all, phishing attacks. On the other hand, a false positive occurs when a legitimate email is incorrectly flagged as phishing. While this can be annoying and can lead to the loss of important information if the email is ignored, the consequences of a false positive are generally not as severe as those of a false negative. Therefore, in the fight against phishing, it is preferable to have more false positives than false negatives. However, the ultimate goal is to minimize both to improve the accuracy of the phishing detection system. In general, the results achieved are acceptable, showing a false positive rate of 0.26 and a false negative rate of 0.10.

4.1 Data Representation Comparison

The approach proposed in this study utilizes a data representation grounded in key phrases and persuasive words. However, recent work by Bountakas and Xenakis [4] introduces a content-based representation that encompasses four distinct categories, as outlined in the related work section. Each of these categories comprises a set of features, totaling 22 in all. These features were employed to represent the dataset presented in this study. Subsequently, we applied three

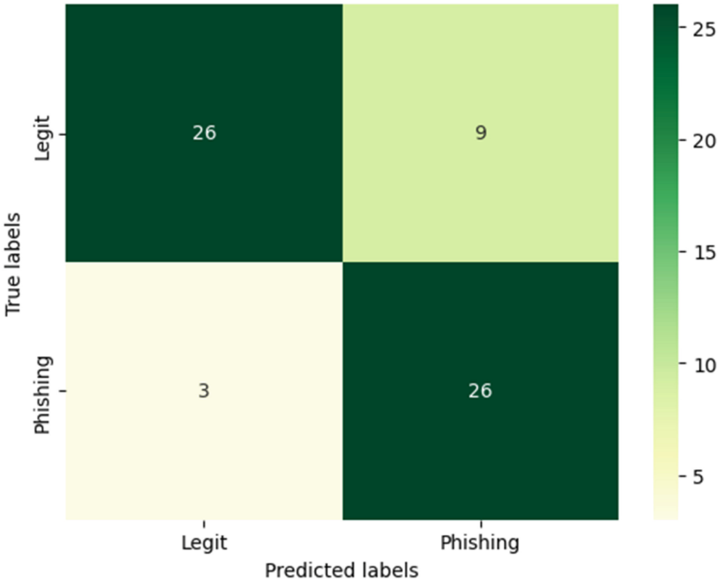


Fig. 6. Confusion matrix obtained using Decision Tree Classifier.

classifiers from different families: Support Vector Machine (SVM), k-Nearest Neighbors (KNN), and Decision Tree Classifier. The resulting outcomes are summarized in Table 1.

Table 1. Comparison of Classifier Performance using different Data Representation

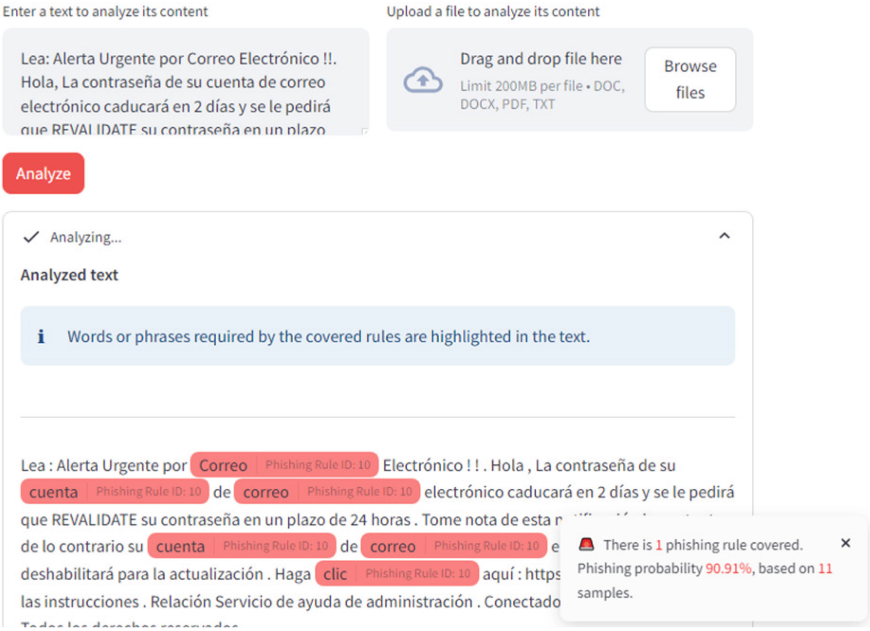
Classifier (Data Representation)	Precision	Recall	F-score	Accuracy
SVM (key phrases and persuasive words)	0.71	0.83	0.76	0.77
SVM (Bountakas and Xenakis [4])	0.70	0.80	0.75	0.75
KNN (key phrases and persuasive words)	0.74	0.84	0.79	0.78
KNN (Bountakas and Xenakis [4])	0.71	0.83	0.76	0.77
Decision Tree (key phrases and persuasive words)	0.74	0.90	0.81	0.81
Decision Tree (Bountakas and Xenakis [4])	0.71	0.86	0.78	0.78

Overall, the results obtained highlight the effectiveness of the data representation proposed in this study for the evaluated classifiers. Specifically, the Decision Tree classifier, when utilizing this data representation, achieved an impressive accuracy of 0.81 and a recall of 0.90. These metrics indicate its strong performance in correctly identifying positive instances. This finding underscores the significance of combining persuasive language cues with key phrases extracted from phishing messages. Such a combination proves valuable for classifiers in effectively distinguishing between legitimate messages and phishing attempts.

4.2 Demo Tool

Once the rules are established, the challenge lies in conveying this knowledge to the average user in a user-friendly manner, making them aware of the need to protect themselves against phishing attacks.

For this purpose, a demo web tool based on Streamlit [25] was created. As depicted in Fig. 7, this tool allows users to interact by inputting text or uploading a file for its content to be analyzed. Upon processing the user-provided information, if any phishing or legitimate rule is met, the tool will alert the user. Additionally, the tool highlights the words within the user's text that triggered the alert. It also informs the user about the type of rule that was triggered and the probability that the email is phishing or legitimate (based on the rule's probability or confidence).


Enter a text to analyze its content
Upload a file to analyze its content
Drag and drop file here
Limit 200MB per file • DOC, DOCX, PDF, TXT
Browse files
Analyze
Analyzing...
Analyzed text

Lea: Alerta Urgente por Correo Electrónico !!. Hola, La contraseña de su cuenta de correo electrónico caducará en 2 días y se le pedirá que REVALIDATE su contraseña en un plazo

Words or phrases required by the covered rules are highlighted in the text.

Lea : Alerta Urgente por Correo Phishing Rule ID: 10 Electrónico ! ! . Hola , La contraseña de su cuenta Phishing Rule ID: 10 de correo Phishing Rule ID: 10 electrónico caducará en 2 días y se le pedirá que REVALIDATE su contraseña en un plazo de 24 horas . Tome nota de esta r
de lo contrario su cuenta Phishing Rule ID: 10 de correo Phishing Rule ID: 10 e
deshabilitará para la actualización . Haga clic Phishing Rule ID: 10 aquí : https
las instrucciones . Relación Servicio de ayuda de administración . Conectado
Todos los derechos reservados

There is 1 phishing rule covered. Phishing probability 90.91%, based on 11 samples.

Fig. 7. Demo web tool created to interact with users.

Although it is currently only a demo tool, work is underway to develop this concept further and enable other options to make it more appealing to users and subtly contribute to their cybersecurity culture.

5 Conclusions

Based on the evaluated data set, this study developed an effective method for detecting phishing attempts in Spanish emails. The results obtained have shown that our approach has an accuracy of 81%, indicating an acceptable accuracy in identifying phishing emails. Such accuracy rate provides significant protection to Spanish email users, minimizing the risk of falling into phishing traps.

The study findings emphasize the importance of leveraging persuasive language cues alongside key phrases from phishing messages, as demonstrated by the Decision Tree classifier remarkable accuracy and recall. This integrated approach holds promise for robustly distinguishing between legitimate messages and phishing attempts.

In addition, as part of this work, we have created a demonstration tool that allows users to interactively evaluate a text and determine its likelihood of being a phishing attempt. This tool not only serves as a practical application of our research but also provides users with a tangible way to understand and apply our findings.

In summary, our work has not only proven to be effective in detecting phishing in Spanish emails, but also has the potential to be adapted to other languages. The proof-of-concept conducted showed promising results, but there is still a lot of development work to be done to create a high-accurate detection tool. The gradual incorporation of new phishing emails to the data set will allow adding more useful information for the construction of new rules that help to improve the results achieved so far. Our ultimate goal is to contribute to a safer and more reliable digital environment for all.

References

1. Alhogail, A., Alsabih, A.: Applying machine learning and natural language processing to detect phishing email. Comput. Secur. **110**, 102414 (2021)
2. Ariyadasa, S., Fernando, S., Fernando, S.: Detecting phishing attacks using a combined model of LSTM and CNN. Int. J. Adv. Appl. Sci **7**(7), 56–67 (2020)
3. Banu, R., Anand, M., Kamath, A., Ashika, S., Ujwala, H., Harshitha, S.: Detecting phishing attacks using natural language processing and machine learning. In: 2019 International Conference on Intelligent Computing and Control Systems (ICCS), pp. 1210–1214. IEEE (2019)
4. Bountakas, P., Xenakis, C.: Helphed: hybrid ensemble learning phishing email detection. J. Netw. Comput. Appl. **210**, 103545 (2023)
5. Boussougou, M.K.M., Jin, S., Chang, D., Park, D.J.: Korean voice phishing text classification performance analysis using machine learning techniques. In: Proceedings of the Korea Information Processing Society Conference, pp. 297–299. Korea Information Processing Society (2021)
6. Bozkir, A.S., Aydos, M.: LogoSense: a companion hog based logo detection scheme for phishing web page and e-mail brand recognition. Comput. Secur. **95**, 101855 (2020)
7. Bustio-Martínez, L., Álvarez-Carmona, M.A., Herrera-Semenets, V., Feregrino-Uribe, C., Cumplido, R.: A lightweight data representation for phishing URLs detection in IoT environments. Inf. Sci. **603**, 42–59 (2022)

8. Bustio-Martínez, L., et al.: Towards automatic principles of persuasion detection using machine learning approach. In: Hernández Heredia, Y., Milián Núñez, V., Ruiz Shulcloper, J. (eds.) IWAIPR 2023. LNCS, vol. 14335, pp. 155–166. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-49552-6_14
9. Cohen, W.W.: Fast effective rule induction. In: Machine Learning Proceedings 1995, pp. 115–123. Elsevier (1995)
10. Dilhara, B.: Phishing URL detection: a novel hybrid approach using long short-term memory and gated recurrent units. Int. J. Compu. Appl. **975**, 8887 (2021)
11. Europol: Phishing gang behind several million euros worth of losses busted in Belgium and the Netherlands (2022). https://www.europol.europa.eu/media-press/newsroom/news/phishing-gang-behind-several-million-euros-worth-of-losses-busted-in-belgium-and-netherlands. Accessed 31 Jan 2024
12. Fürnkranz, J., Widmer, G.: Incremental reduced error pruning. In: Machine Learning Proceedings 1994, pp. 70–77. Elsevier (1994)
13. Herrera-Semenets, V., Bustio-Martínez, L., Hernández-León, R., van den Berg, J.: A multi-measure feature selection algorithm for efficacious intrusion detection. Knowl.-Based Syst. **227**, 107264 (2021)
14. Hiransha, M., Unnithan, N.A., Vinayakumar, R., Soman, K., Verma, A.: Deep learning based phishing e-mail detection. In: Proceedings of 1st AntiPhishing Shared Pilot 4th ACM International Workshop Security Privacy Analysis (IWSPA), pp. 1–5. Tempe, AZ, USA (2018)
15. Lee, J., Xin, Z., See, M.N.P., Sabharwal, K., Apruzzese, G., Divakaran, D.M.: Attacking logo-based phishing website detectors with adversarial perturbations. In: Tsudik, G., Conti, M., Liang, K., Smaragdakis, G. (eds.) ESORICS 2023. LNCS, vol. 14346, pp. 162–182. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-51479-1_9
16. Lee, M., Park, E.: Real-time Korean voice phishing detection based on machine learning approaches. J. Ambient. Intell. Humaniz. Comput. **14**(7), 8173–8184 (2023)
17. Moussavou Boussougou, M.K., Park, D.J.: Attention-based 1D CNN-BiLSTM hybrid model enhanced with fasttext word embedding for Korean voice phishing detection. Mathematics **11**(14), 3217 (2023)
18. Naqvi, B., Perova, K., Farooq, A., Makhdoom, I., Oyedeji, S., Porras, J.: Mitigation strategies against the phishing attacks: a systematic literature review. Comput. Secur. 103387 (2023)
19. Pérez-Guadarramas, Y., Simón-Cuevas, A., Romero, F.P., Olivas, J.A.: Topic modeling based on OWA aggregation to improve the semantic focusing on relevant information extraction problems. In: Rivera, G., Cruz-Reyes, L., Dorronsoro, B., Rosete, A. (eds.) Data Analytics and Computational Intelligence: Novel Models, Algorithms and Applications. Studies in Big Data, vol. 132, pp. 17–42. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-38325-0_2
20. Proofpoint: 2023 state of the phish: Europe and the middle east (2023). https://www.proofpoint.com/uk/resources/threat-reports/state-of-phish. Accessed 31 Jan 2024
21. Ra, V., HBa, B.G., Ma, A.K., KPa, S., Poornachandran, P., Verma, A.: Deepanti-phishnet: applying deep neural networks for phishing email detection. In: Proceedings of 1st AntiPhishing Shared Pilot 4th ACM Int. Workshop Security Privacy Analysis (IWSPA), pp. 1–11. Tempe, AZ, USA (2018)
22. Sahingoz, O.K., Buber, E., Demir, O., Diri, B.: Machine learning based phishing detection from URLs. Expert Syst. Appl. **117**, 345–357 (2019)

23. sklearn: Decisiontreeclassifier (2024). https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html. Accessed 5 Feb 2024
24. Statista: Los idiomas mas hablados en el mundo en 2023 (2024). https://es.statista.com/estadisticas/635631/los-idiomas-mas-hablados-en-el-mundo/. Accessed 31 Jan 2024
25. Streamlit: Api reference (2024). https://docs.streamlit.io/library/api-reference. Accessed 5 Feb 2024
26. Thakur, K., Ali, M.L., Obaidat, M.A., Kamruzzaman, A.: A systematic review on deep-learning-based phishing email detection. Electronics **12**(21), 4545 (2023)
27. Vazhayil, A., Vinayakumar, R., Soman, K.: Comparative study of the detection of malicious URLs using shallow and deep networks. In: 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–6. IEEE (2018)
28. Wang, M., Zang, X., Cao, J., Zhang, B., Li, S.: Phishhunter: detecting camouflaged IDN-based phishing attacks via Siamese neural network. Comput. Secur. **138**, 103668 (2024)
29. Wei, W., Ke, Q., Nowak, J., Korytkowski, M., Scherer, R., Woźniak, M.: Accurate and fast URL phishing detector: a convolutional neural network approach. Comput. Netw. **178**, 107275 (2020)
30. Yang, J., Lee, C., Kim, S.: Development and utilization of voice phishing prevention service through koBERT-based voice call analysis. KIISE Trans. Comput. Pract **29**, 205–213 (2023)
31. Zhang, Q., Bu, Y., Chen, B., Zhang, S., Lu, X.: Research on phishing webpage detection technology based on CNN-BiLSTM algorithm. J. Phys. Conf. Ser. **1738**, 012131 (2021). IOP Publishing