An Application for Predicting Phishing Attacks: A case of Implementing a Support Vector Machine Learning Model

Emmanuel Song Shombot, Gilles Dusserre, Robert Bestak, Nasir Baba Ahmed

Cyber Security and Applications 2 (2024) 100036

Contents lists available at ScienceDirect

Cyber Security and Applications

journal homepage: http://www.keaipublishing.com/en/journals/cyber-security-and-applications/

An application for predicting phishing attacks: A case of implementing a support vector machine learning model

Emmanuel Song Shombot [a,*], Gilles Dusserre [a], Robert Bestak [b], Nasir Baba Ahmed [c]

[a] Laboratory for the Science of Risks (LSR), IMT Mines Ales, 6 Av. de Clavières, Alès 30100, France
[b] Czech Technical University in Prague, Technicka 2, 6, Prague 16627, Czech Republic
[c] IMT Mines Ales, 6 Av. de Clavières, Alès 30100, France

ARTICLE INFO

Keywords:
Phishing
SVM
Cybersecurity
Healthcare
Machine learning
AI
Cyberattack

ABSTRACT

The imminent threat that phishing websites poses is a major concern for internet users worldwide. These fraudulent websites are crafted by cyber attackers to appear trustworthy and deceive vulnerable users into divulging confidential data like medical health records, credit card details, passwords, and Personal Identifiable information (PII). To bait their victims, cybercriminals employ tactics such as social engineering, spear-phishing attacks, and email phishing scams. As a result, unsuspecting individuals may be enticed to visit these websites, putting their sensitive information at risk. This work presents an application designed to predict phishing attacks after comparing polynomial and radial basis function of support vector machine (SVM). The proposed application leverages a dataset of known legitimate, suspicious and phishing attacks stored in a database and employs an SVM algorithm for classification based on user input. The application provides a user-friendly graphical user interface (GUI) that allows reporting of new phishing incidents based on the features that have strong relationship in determining if a website is phishing or not. The proposed application utilizes the inherent scalability of database technology to support record expansion whenever there is an instance of a user initiating phishing prediction thereby, making it suitable for use in a wide range of organizational settings.

1. Introduction

The world experienced a paradigm shift in the *modus operandi* of cybercriminals since the COVID-19 outbreak as more than 150 countries experienced partial or complete movement restriction alongside significant alteration in the method in which economic activities are conducted [1,2]. Cybercrime is an illegal action aimed at computer systems or networks, encompassing a wide spectrum of potentially criminal activities [3]. Cyberattack could be directed at vulnerable computer networks or, it can rely on the victim's implicit participation in the attacker's criminal scheme for the activity to be successful as in the case of social engineering attacks. As defined by [4] and [5], social engineering attack is the art of psychologically manipulating an individual through persuasion to reveal sensitive or confidential information. Amongst the popular types of social engineering attacks such as trojan horse, shoulder surfing and dumpster diving [6], phishing stands out as the most frequently employed technique [2,3].

Phishing is a method of cyberattack whereby cyber criminals attempt to get hold of people's personal identifiable information by misleading them using psychological trickery [7]. Other authors as Merwe et al., [8] considers phishing as "a fraudulent activity that involves the creation of a replica of an existing web page to fool a user into submitting personal, financial, or password data." Although, definitions of phishing attacks can be fragmented by focusing on social engineering aspect and theft of PII, a more robust definition by Alkhalil et al., [9] suggest "phishing as a socio-technical attack, in which the attacker targets specific valuables by exploiting an existing vulnerability to pass a specific threat via a selected medium into the victim's system, utilizing social engineering tricks or some other techniques to convince the victim into taking a specific action that causes various types of damages." These threats can range from malicious web links, attachments and fraudulent data entry forms. The criticality of this attack method cannot be over emphasized as Sánchez-Paniagua [10] mentions that, phishing is the most challenging social engineering attack to curb, due to the large number of people currently engaging in online activities and that makes it certainly more challenging to detect and prevent. A recent Proofpoint study reported by Techopedia in 2023, reveals that a staggering 83 % of companies fall victim to phishing attacks annually, highlighting the pervasive nature of this cyber threat. The alarming trend is further underscored by a substantial 345 % surge in unique phishing sites observed between 2020 and 2021. The FBI's Internet Crime Complaint Center (IC3) reports a significant escalation in phishing incidents, with a staggering

* Corresponding author.
E-mail address: emmanuel.song@mines-ales.fr (E.S. Shombot).

https://doi.org/10.1016/j.csa.2024.100036
Received 21 August 2023; Received in revised form 16 November 2023; Accepted 8 January 2024
Available online 17 January 2024

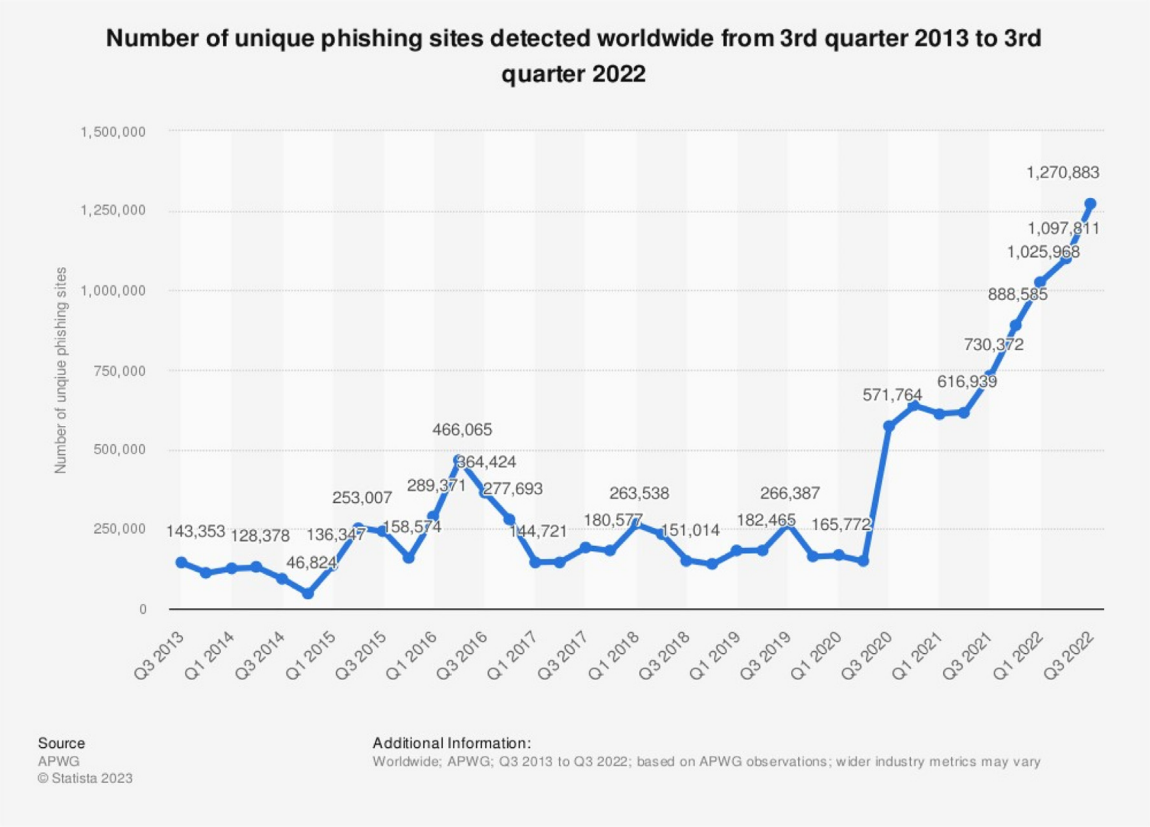

Fig. 1. Number of phishing websites from 2013 to 2022. [18].

800,944 reports and losses exceeding $10.3 billion in 2022 alone. The financial repercussions are substantial, as each phishing attack carries an average cost of $4.91 million for corporations. These statistics underscore the pressing need for robust cybersecurity measures and heightened awareness to counter the growing sophistication and prevalence of phishing attacks in the digital landscape [11]. Phishing attacks mostly include fraudulent emails [12], websites [13], phone calls [14], text messages [15] that appear to be from legitimate sources such as banks, social media platforms, or government agencies and the successful outcome in some cases is the installation of dangerous malware [16]. Due to the widespread use of online transactions and services, phishing attacks have become a major concern for individuals and organizations alike. As shown in Fig. 1, it is quite clear that the number of unique phishing sites detected worldwide from 3rd quarter 2013 to 3rd quarter 2022 has been on the rise, with a dramatic surge in the year 2020, which most likely can be attributed to the COVID-19 pandemic that caused some major disorientation in the general conduct of economic activities.

This work emphasis is primarily on using website features to predict phishing websites based on SVM machine learning algorithm because as mentioned by Cui et al. 2020 [17], the features required to detect phishing attacks are different depending on the attack vector used. Therefore, the key contribution of this work to knowledge is;

1. Implement a machine learning algorithm that will predict phishing websites based on their features.
2. Develop a web application that will assimilate the phishing detection algorithm in an interactive way and provide a platform for users to make online predictions based on the identified features.
3. Capture each user prediction interaction session as a new record in a database for the purpose of expanding the record in the database and improving the performance of the machine learning algorithm.

2. Literature review

Generally, the practice of using ML in cybersecurity is still in its infancy or experimental stages, demonstrating a substantial gap between research and practice. As a result, the use of machine learning in cybersecurity is presented in a very disjointed manner, which makes it difficult to deploy in practice [19,20]. The promise of artificial intelligence and machine learning as presented in most literature in terms of what they can achieve in cybersecurity can at best be considered speculative [21], largely because AI and ML are exclusively data-driven, and currently, the availability of such data is lacking or too specific to a given use case that it cannot be reproduced for other identified contexts [22]. In terms of the practical application of AI and ML as implemented in most use cases presented in some of the most recent literature [23], a clearer picture can be painted as to the extent of research and what aspects of cybersecurity can be achieved by leveraging AI and ML methods [24]. As far as the three successful applications of ML in cybersecurity as articulated by Apruzzese et al. [19], which include machine Learning in network intrusion detection, Machine learning in malware detection, and Machine learning in phishing detection, this work focuses on phishing detection using machine learning and this section explores all the relevant literature on the subject matter.

2.1. Systematic literature review of phishing

A systematic literature survey was conducted by Safi and Singh, 2023 [25], intended to provide a comprehensive analysis of the techniques used in detecting phishing websites. Their work uncovers that despite the extensive literature search, a comprehensive overview of all the significant approaches employed in this domain was inadequate. Additionally, there was a lack of a systematic resource that aggregates the methodologies, data sets, and algorithms utilized in phishing website de-

tection. Thus, there was a requirement for an authorized review to study this field and provide a comprehensive summary. The main contribution of Safi and Singh, 2023 [25] is to firstly unearth the most effective techniques for detecting phishing websites, so that security managers can effortlessly choose the most effective method from a range of anti-phishing approaches for their security systems and secondly, provide a good systematic literature review paper that will succinctly capture the current state of techniques, data sets and algorithms used to deal with phishing problem.

Another review paper by Qabajeh et al., 2018 [26], titled recent review of conventional vs. automated cybersecurity anti-phishing techniques provides another dimension of review. Their work explores Integrating a classification system with intelligent machine learning technology in the browser as a promising anti-phishing approach that detects and alerts users to phishing activities. Their paper is design to be mostly encompassing as it reviews and analyzes legal, training, educational, and intelligent anti-phishing approaches and highlights their similarities, differences, positive and negative aspects from user and performance perspectives. The study also identifies ways to combat phishing through intelligent and conventional methods, making it beneficial for computer security experts, web security researchers, and business owners.

Another comprehensive survey study was carried out by Basit et al. 2021 [27] where the authors comprehensively explored AI-enabled phishing attacks detection techniques and extrapolated that, most phishing attacks detection methods fall into four categories. Deep learning, which is the latest progress in deep learning methodologies proposes that the categorization of phishing websites using deep neural networks (NN) could surpass the conventional machine learning (ML) algorithms. Nonetheless, the outcomes of employing deep NN significantly rely on the configuration of various learning parameters [28]. Secondly, machine learning method, considered to be popular because it appears that most phishing attacks types are classification problems. The degree of accuracy is relatively high when using this detection method but that depends largely on the dataset and the features therein [29,30]. Thirdly, Scenario-based phishing attack detection method, that is predicated on different scenarios however, these scenarios yield different outcomes based on methods used [27]. Some examples of this method include Begum and Badugu, 2019 [31] that relies on the consolidation of techniques such as Machine Learning (ML) based approaches, Non-machine Learning-based approaches, Neural Network-based approaches, and Behavior-based detection approaches for the detection of phishing attacks. Other authors such as Fatima et al., 2019 [32] presented PhishI for security training based on gaming and Chiew et al., 2018 [13] focused on phishing attack detection based on their features, medium and vectors. Lastly, Hybrid learning (HL) based phishing attack detection suggest the most recent future direction for phishing attack detection which could be based on leveraging more than one machine learning model as in the case of Pandey et al., 2020 [33] where they proposed random forest and support vector machine algorithm as a hybrid model for phishing detection.

2.2. Conventional phishing detection methods

According to Hong J. 2012, there are three main ways to combat phishing attacks [16]: by implementing invisible protections that require no action from the user, by creating better user interfaces, and by providing effective training[16]. There are currently over 500 toolkits available for phishing attack [34], some of which are designed to trick the phisher into providing false information. Criminals and security professionals are engaged in a constant competition to outsmart each other [16].

Phishers use various techniques, such as fast flux, which involves using a pool of proxies and domain names to hide the location of the phishing website [16]. This technique can extend the average lifespan of a phishing website to 196 hours, compared to the average of 62 hours before the location is taken down [35]. Due to the growing threat of phishing and the negative impact it has on the economy and reputation of businesses, it is critical to have effective countermeasures in place, such as filters, machine learning, blacklists, and active and passive indicators that alert users. While the combination of the aforementioned countermeasures with staff training enhances the ability to deal with phishing, it's essential to acknowledge that training users may not always be effective. This is often due to factors such as low motivation or unwillingness to engage with training materials [16].

In a similar study conducted by Minocha and Singh, 2022 [36], they identify that there are two categories of automatic detection techniques for phishing sites: (i) list-based systems, and (ii) machine learning-based techniques. List-based systems use safelists and denylists to categorize websites as legitimate or phishing, respectively. Additionally, the article points out that the traditional methods of detection only provide around 20 % success.

Tan et al. [37] propose a phishing detection method that is based on comparing the actual and target identities of a webpage. The proposed method, PhishWHO, comprises three phases:

1. Extracting identity keywords from the website's textual content, employing a unique weighted URL tokens system based on the N-gram model.
2. Identifying the target domain name through a search engine, then selecting the domain based on identity-relevant characteristics.
3. Proposing a 3-tier identity matching system to assess the authenticity of the queried webpage.

Data mining approach as proposed by Abdelhamid et al., [38], considers phishing a typical classification problem and the objective of the classification task is to categorize a new website into one of the predefined classes, such as phishing, legitimate or suspicious. After a website is loaded on the browser, a set of feature values is extracted, which play a crucial role in determining the website type. By utilizing the rules derived from historical data. Also, their work alongside that of [39,40] and [41] based on associative classification which is a data mining technique that combines classification and association rule mining.

Content based approach is another method of detection articulated by Nguyen et al., [42], Zhang et al., [43], and Jha et al., [44]. This method classifies websites as either phishing or non-phishing. This approach relies on analyzing the contents of the site to determine its classification. The "Term Frequency/Inverse Document Frequency" (TF-IDF) algorithm is commonly used for this type of content analysis. Zhang et al., used a similar content-based approach in their research, which they called CANTINA. Their results showed that this technique had a high accuracy rate of 97 % in identifying phishing websites. However, to reduce false positives, heuristics were applied, resulting in a decrease in accuracy to around 90 %. Other sources such as [45–48] also discussed and implemented other variants of content-based approach to phishing detection to corroborate the works of Zhang et al. [43].

Various approaches have been explored in the literature to enhance phishing detection [49], each accompanied by its own set of drawbacks. One method involves specifying weights for words extracted from URLs and HTML contents, focusing on elements like brand names, with a dependency on a third-party server, Yahoo Search, resulting in an accuracy rate of 98.20 % [50]. However, a drawback of this approach is its reliance on an external server and over dependence on textual content. Another strategy utilizes logo image analysis to identify web page authenticity, matching real and fake webpages, but with a dependency on Google Image Search and an accuracy rate of 93.40 % [51]. The drawback here is the reliance on a third-party server and the exclusivity of this method to only use images. In another method, the use of URL heuristics and website rank for detection is implemented, but the drawback lies in the time-consuming process of feature extraction and website rank examination, achieving an accuracy rate of 97.16 % [52].

Rao and Ali implemented an advanced version of this technique using a desktop application called "Phish Shield." They used novel heuris-

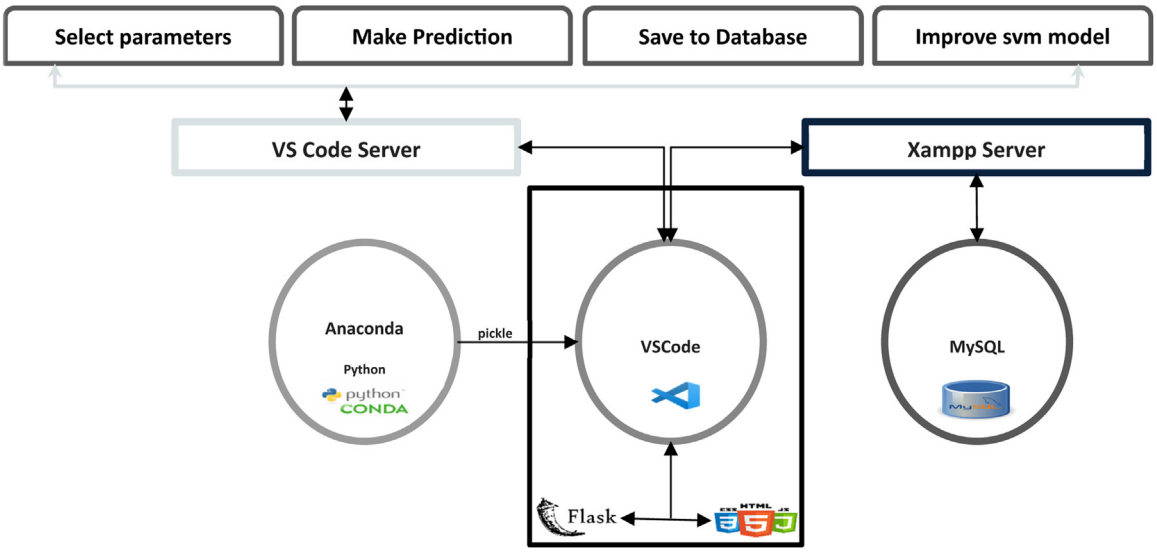

Fig. 2. Visual abstract of proposed system.

tics based on the URL to reduce false positives to 0.035 % and achieve an accuracy of 96.57 %. Their approach included null footer links, maximum frequency domains, copyrights, and whitelists for the detection process. However, their approach had a limitation in terms of response time, which could be improved with newer methodologies such as genetic algorithms and neural networks.

2.3. Machine learning approach

The work of Alani and Tawfik, 2022 [15], pivots on building a machine-learning-based phishing detection system using only the URL as they argue that their approach provides better network protection by reducing the attack surface. They also applied recursive feature elimination (RFE) which is a very useful feature selection method to reduce the number of features to the most important and critical features. Their work leveraged a pipeline for five machine learning classifiers of random forest, logistic regression, decision tree, gaussian naïve bayes and multi-layer perceptron using a 75 % training and 25 % testing set. Overall, random forest performed better than the other models they tested. Finally, in contrast to other locally hosted phishing solutions, their system can be deployed to the cloud as an API to be integrated as a browser plugin.

Sahoo et al. [53] discuss various malicious attacks, as well as different types of machine learning and features for detecting malicious URLs. The paper primarily focuses on the identification of features used for classifying malicious websites, grouped into five categories, and highlights the design and limitations of some of these features. The authors also provide examples of machine learning algorithms and their application in detecting malicious websites.

Ensemble learning classification is another method popularized for phishing detection which is based on using multiple classifier or algorithms to solve a classification problem [54]. The work carried out by Al-Sarem et al., 2021 [55], proposed an optimized stacking ensemble model for phishing websites detection. Their approach includes three stages of training, ranking and testing. The classifiers, namely random forests, AdaBoost, XGBoost, Bagging, GradientBoost, and LightGBM, were initially trained without utilizing any optimization method. Subsequently, the genetic algorithm was employed to optimize these classifiers by determining the most favorable parameter values for various ensemble models. This process enabled the selection of the optimal parameters for these classifiers. Another work by Abawajy and Kelarev, 2012 [56] used a multi-tier ensemble construction of classifiers for phishing however, their work only focused on email detection and filtering. Similarly, Bountakas and Xenakis, 2023 [57], proposed hybrid ensemble learning PHishing email detection based on stacking and soft voting.

3. Proposed system

The development of the proposed system starts by using a dataset of website characteristics labeled as either legitimate, suspicious or phishing. Based on the datapoints, the SVM algorithm is utilized and optimized using polynomial or radial basis function to determine which kernel provides better model accuracy with minimal errors. The trained model is then used to create a web application that runs on a server. This application takes features of a website as input and outputs a classification label based on the trained model. To ensure that the SVM model remains accurate over time, a database is created to store new user interactions with websites. These interactions can be analyzed and fed back into the SVM model to improve its accuracy. This continuous improvement process ensures that the model remains up-to-date with the latest trends and techniques used by phishing websites. In summary, creating a machine learning model to identify and differentiate between legitimate, suspicious, and phishing websites involves training the model using SVM, creating a web application to implement the model, creating a database to store user interactions, and continuously improving the model based on new data. By following these steps, the accuracy and reliability of the model can be maximized, and users can be better protected against phishing attacks Fig. 2.

4. Methodology

As depicted in Fig. 3, the methodology for this paper begins by identifying the phishing dataset by Abdelhamid et al., [38], and preprocessing it to assess its suitability for the intended task. Next, we identified the independent features (input variables) and the class label (output variable). Clearly defining these elements is fundamental to the training of a machine learning model. Subsequently, we determined the nature of the phishing problem we aim to solve based on the dataset's characteristics. Utilizing a pair plot library in Python, we identified the relationships between pairs of variables in the dataset. From the outcomes of the pair plot, we established that we are dealing with a classification problem. Therefore, we progressed to identify and test several supervised learning classification algorithms such as Random Forest, gradient boosting, decision trees, and SVM. We then proceeded to build and train the model, selecting SVM as it exhibited the highest accuracy in our case. Additionally, we implemented two optimization kernels of SVM, namely RBF and

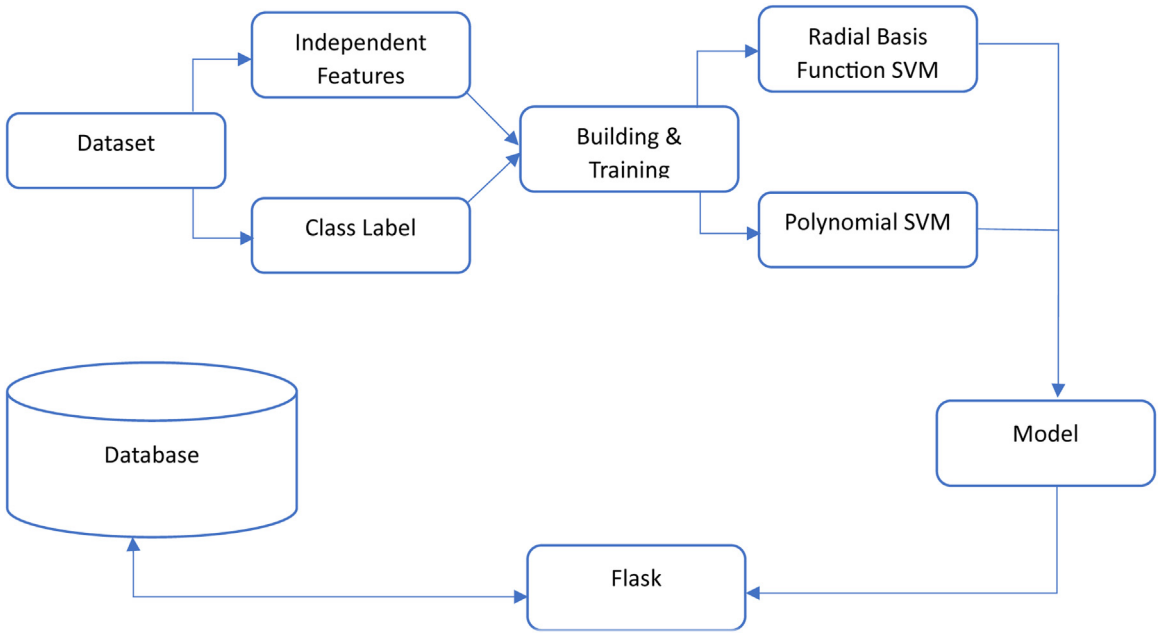

Fig. 3. Model architecture.

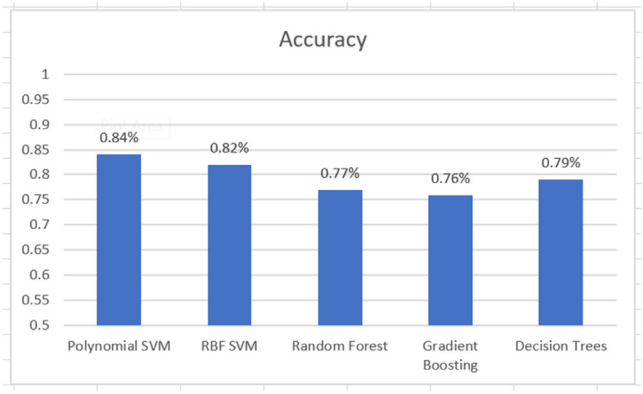

Fig. 4. Accuracy comparison for classification algorithms.

polynomial functions. The next step involves integrating the model into a web framework for deployment. The last crucial stage is incorporating database technology to interact with Flask, particularly to support the expansion of the database.

4.1. Justification for algorithm used

Different algorithms are better suited for different types of data and different problems. So the choice of algorithm used for this work depends on the specific characteristics of the data, the problem we are solving, and the trade-offs between model complexity, interpretability, and computational efficiency [58]. For instance, logistic regression is used in classification problems where the datapoints are linearly separable however, that will not be applicable in this case given that our data is a three-label multiclass classification problem [59]. Therefore, we experimented with classification algorithms that are better suited for multiclass classification problem such as SVM, Random Forest, Gradient boosting and Decision trees. From the results shown in Fig. 4, the SVM model performed relatively better than the other models. There are also theoretical dimensions that are abundantly articulated in the literature on the use of SVMs in solving multiclass classification problems [60–63]. Some of these inherent advantages include;

1. **Handling Nonlinear Relationships:** SVMs are effective in capturing complex nonlinear relationships in data.
2. **Performance in Imbalanced Datasets:** SVMs can handle imbalanced datasets well and are not heavily influenced by the class distribution.
3. **Memory Efficiency:** SVMs are memory-efficient, particularly when dealing with high-dimensional datasets than most of the tree-based models.

4.2. Implementation environment

The hardware and software specifications of the implementation environment used in this work.

1. Laptop HP EliteBook pro, x360 8th Gen, 500GB SSD, Intel Core i7, CPU 2.11GHz, Ram 16GB
2. Anaconda Distribution Version 3 (Jupyter lab – 6.4.12 running on localhost 8888, Python 3.9, pandas, SKlearn, matplotlib)
3. Xampp for windows Version 8.2.0 (Control Panel V3.3.0, Apache, MySQL)
4. VS Code Version 1.77 (Flask, CSS, HTML, Javascript)

4.3. Data source

For the dataset used in the development of the application, this article leverages the work of Abdelhamid et al. [38]. The dataset name is titled “website phishing” made available on the 11/1/2016 with multivariate characteristics, designed for classification tasks and consist of integer-type features [64]. Several features related to legitimate and phishing websites were identified, and a dataset comprising 1353 websites from diverse sources was collected. The Phishtank data archive (www.phishtank.com), a community website that allows users to submit, verify, track and share phishing data, was the source of phishing websites. Legitimate websites were sourced from Yahoo and starting point directories using a PHP web script. The PHP script was integrated with a browser, enabling the authors to collect 548 legitimate websites out of the 1353 total websites. The dataset consisted of 702 phishing URLs and 103 URLs classified as suspicious. When a website is deemed SUSPICIOUS, it indicates that it displays features that are characteristic of both legitimate and phishing websites, implying that the website has both genuine and fraudulent attributes.

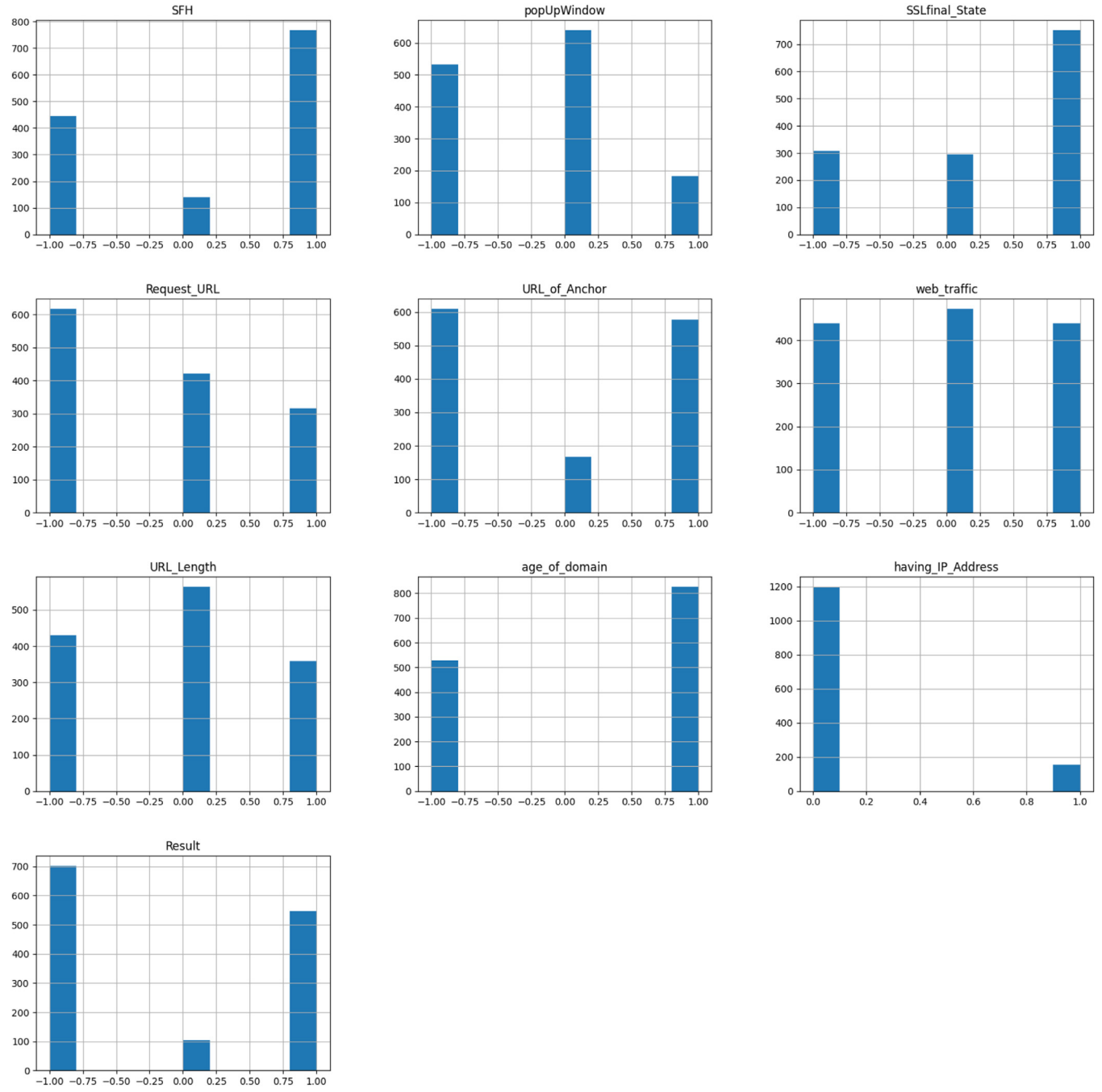

Fig. 5. Feature set distribution.

Furthermore, another impressive thing about the dataset is they applied the Chi-Square measure [65], which is a statistical method used to determine the degree of association or independence between two categorical variables. It involves calculating the difference between observed and expected frequencies of occurrence in a contingency table, and then comparing these differences to their expected values under the assumption of independence. By doing this, it was possible to come up with concise features that are precise and avoid noise in data that may not be necessarily important in making accurate predictions. After the application of the Chi-square test, 9 features had the most correlation with class attribute among the 16 initially identified features. These features are Request URL, Age of Domain, HTTPS and SSL, Website Traffic, Long URL, SFH, Pop-Up window, URL of Anchor, Redirect URL and Using the IP Address.

4.4. Exploratory data analysis

As captured in Fig. 5. The EDA captures the features of the dataset and its distribution across the class label of [-1 = phishing, 0 = suspicious, 1 = legitimate] Fig. 7.

4.5. Support vector machine

Support Vector Machines (SVMs) are a family of generalized linear classification methods used for both classification and regression tasks in supervised learning [67]. They have a special property of simultaneously minimizing the empirical classification error and maximizing the geometric margin, which has earned them the nickname of Maximum Margin Classifiers. SVMs are based on the Structural Risk Minimization

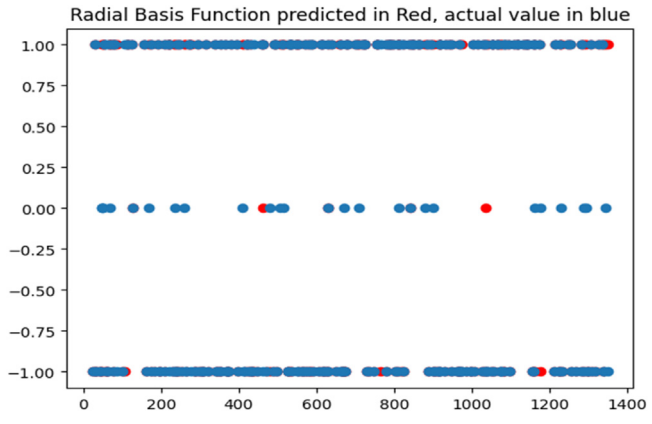

Fig. 6a. Result of radial basis function implementation.

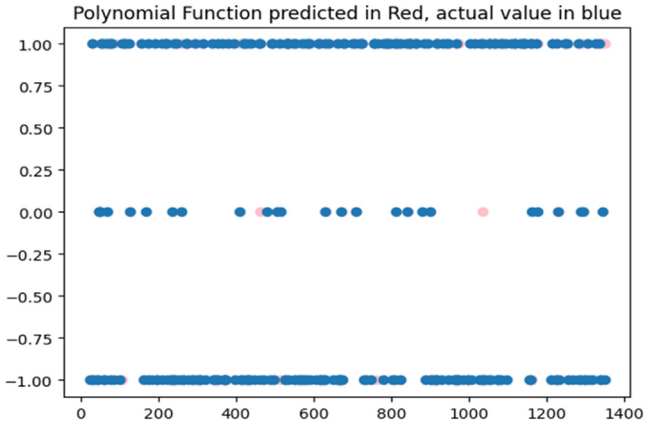

Fig. 6b. Result of polynomial function implementation.

(SRM) principle and work by mapping the input vectors into a higher dimensional space where a maximal separating hyperplane is constructed. Two parallel hyperplanes are constructed on each side of the separating hyperplane to separate the data. The hyperplane that maximizes the distance between these parallel hyperplanes is chosen as the separating hyperplane. It is assumed that a larger margin or distance between the parallel hyperplanes leads to better generalization error of the classifier [67,68].

In this work, the SVM algorithm utilizes a hyperplane to effectively distinguish between different data elements and classify them as phishing, legitimate or suspicious based on the dataset. By separating the features, the hyperplane ensures the best separation of data. SVM is then mapped into the same space, and it predicts the category based on which side of the gap the point or input falls on. Furthermore, in implementing SVM, two kernels of Radial Basis function and Polynomial were tested on the dataset to determine which one works optimally with the dataset.

4.5.1. Radial basis function kernel

This is the kernel function used in the radial basis function (RBF) kernel of support vector machines. The mathematical equation for the RBF kernel function is [68]:

$$K(xi, xj) = \exp(-\gamma||xi - xj||^2)$$

where xi and xj are input vectors, $||.||$ denotes the Euclidean distance between xi and xj, and γ is a hyperparameter that controls the width of the kernel.

This kernel function maps the input data into a higher-dimensional feature space where it becomes separable by a linear decision boundary. The RBF kernel is widely used in support vector machines due to its ability to handle complex, nonlinear decision boundaries in the data [68].

4.5.2. Polynomial kernel

A polynomial kernel is a type of kernel that can be used in SVMs. It is defined as follows [68]:

$$K(xi, xj) = (\gamma xi^T xj + r)^d$$

This is the equation for the polynomial kernel function used in support vector machines (SVMs). The kernel function calculates the similarity between two data points, xi and xj, by computing the dot product of their feature vectors and raising it to the power of d, while adding a

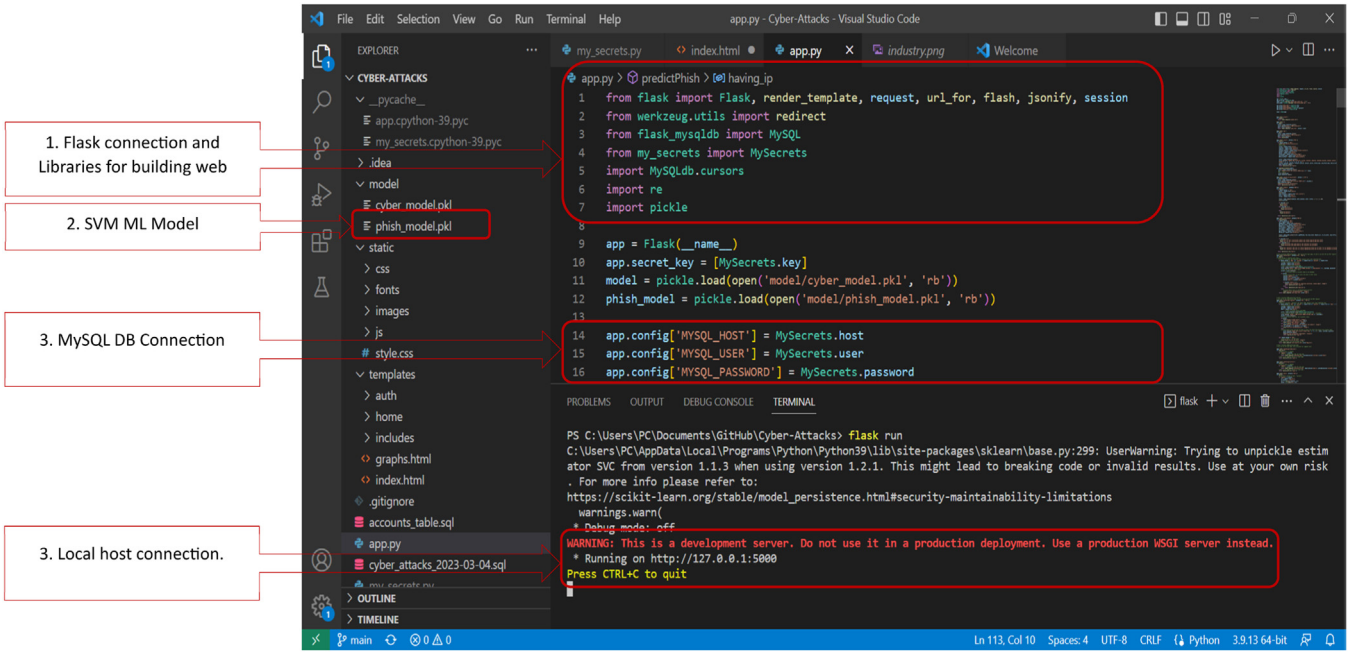

Fig. 7. Code snippet.

Table 1
Features and definition.

SN.	Feature	Description
1.	Server Form Handler (SFH)	When user submits information, it is transferred to a server for processing, typically on the same domain, but phishers may use an empty server form handler or transfer the information to a different domain.
2.	Pop-up Windows	High prevalence of pop-up windows pre-empting users to enter their personal identifiable information is more likely to be associated with phishing sites.
3.	SSL Certificate	HTTPS protocol presence indicates a legitimate website, but phishers may use fake HTTPS, hence verify HTTPS by trusted issuers (e.g., GeoTrust, GoDaddy, VeriSign).
4.	Web Redirect	Phishers often use link redirection to deceive users into submitting their information to a fraudulent site, making it difficult for users to detect the real link they are being directed to.
5.	@ Symbol	The "@" symbol leads the browser to ignore everything prior it and redirects the user to the link typed after it.
6.	Web Traffic	Phishing websites have low web traffic and short life, whereas legitimate websites have high traffic and lower rank, typically less than or equal to 150,000 according to Alexadatabase.
7.	Long URL	Phishers may hide parts of the URL to redirect user information or upload pages to suspicious domains, with no reliable length to distinguish phishing from legitimate URLs, but a length greater than 54 characters may indicate a phishing URL [66]
8.	Age of domain	Websites with a duration of less than 1 year of online presence may be deemed risky.
9.	IP address	Using an IP address in the domain name of the URL is an indicator someone is trying to access the personal information. This trick involves links that may begin with an IP address that most companies do not commonly use any more.

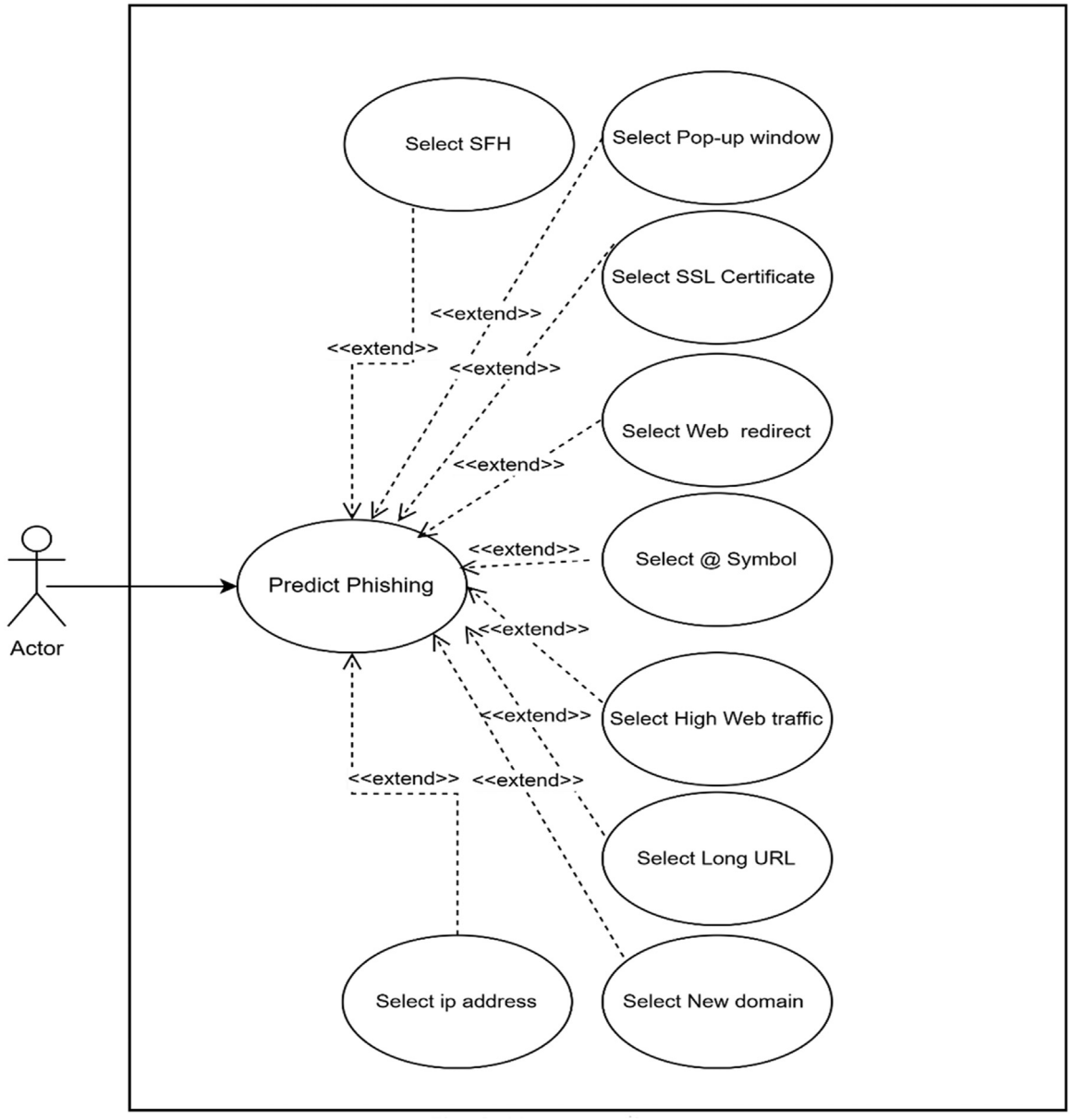

Fig. 8. Use case diagram.

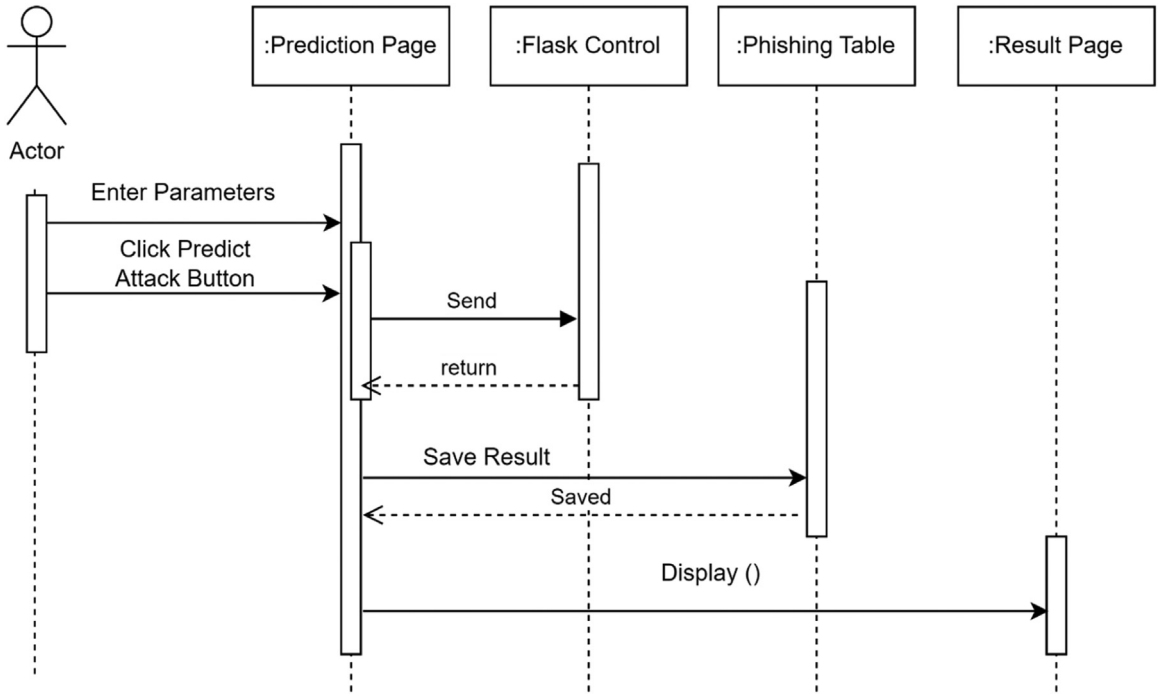

Fig. 9. Sequence diagram.

constant r. The value of γ determines the width of the kernel and affects the smoothness of the decision boundary. The polynomial kernel is used to transform the input space to a higher dimensional space to achieve a better separation of data points [68].

4.6. Web application implementation

To build the web application for prediction, SVM model was used and two kernels of radial basis function and polynomial function were tested. Given that the polynomial function performed better as a kernel, the model was saved as a pickle file and imported into Visual studio code which is a lightweight but powerful source code editor running on a local machine. Using Flask as a Python web framework that provides useful tools and features that make creating web applications easier, it was possible to feed input data as a python variable for output prediction appearing as a HTML file. Flask is very important in this web application implementation because, it allows for the integration of ML models created using sklearn library in python into an interactive application for users. Since this is a web application, javascript was used as the programming language, HTML was used to structure the application and CSS to design and layout the webpage. Furthermore, to ensure the continuous growth of the data, a database was designed using MySQL database running on the local machine and it uses the port number "3306". The phishing table in the database captures all instances of phishing prediction initiated by a user and saving the parameters into the database. To get the application up and running, the local host service was utilized at url "http://127.0.0.1:5000/". This setup is what allows the web application to run and carry out the functionalities it is intended to perform Table 1.

4.7. User interaction

To visualize the interactive behavior of the system, a use case diagram is used as depicted in Fig. 8. Which is very important for modelling the dynamic behavior of a system. Another model used is the sequence diagram as captured in Fig 9. The sequence diagram shows how the dynamic scenario and message passing between multiple system objects when initiated by an actor or user. Finally, the appearance of a graphical user interface (GUI) as implemented for the user to interact with to predict a phishing website attack is captured in Fig. 10.

5. Results and future scope

From the results gotten at the first stage of this work, which is developing the machine learning model in the Anaconda environment using Python and relevant libraries such as Pandas for data manipulation, Matplotlib for visualization, and Sklearn as the algorithm for developing the SVM model, the outcome of the model shows that the polynomial function performed better with 84.5 % accuracy, as captured in Fig. 6a., where the actual value is predicted in blue and the polynomial function in red. In contrast, the radial basis function had an accuracy score of 82.6 %, as captured in Fig. 6b, where the actual value is predicted in blue and the radial basis function in red.

The dataset also provides nine features with succinct cardinality that enable feeding the model finely into the web application framework using Flask. In machine learning, datasets with fewer but high-quality features are better in producing more accurate results devoid of unnecessary complexities. Even though 84.5 % is a good accuracy score for the phishing prediction in comparison with the works of [69] with an accuracy score of 77 %, it is also worth mentioning that solving a multi-class classification problem in contrast to a binary classification problem presents a layer of complexity that impacts the accuracy score of the model. Also, the model is designed in such a manner that it will self-improve itself based on the prediction information inputted by the user. This is essentially where the use of the database is paramount to support new data and feed it back to the model to self-correct in a simultaneous fashion.

In comparison with other works done on phishing detection using machine learning, the scope covered by most authors [70–78] is developing the machine learning model using a programming language like Python without paying attention to its practical application from a user's point of view. The results of this work are a clear step beyond what is commonly obtainable because it provides an interactive means, as captured in Fig. 10, where input data can be fed by the user in anticipation

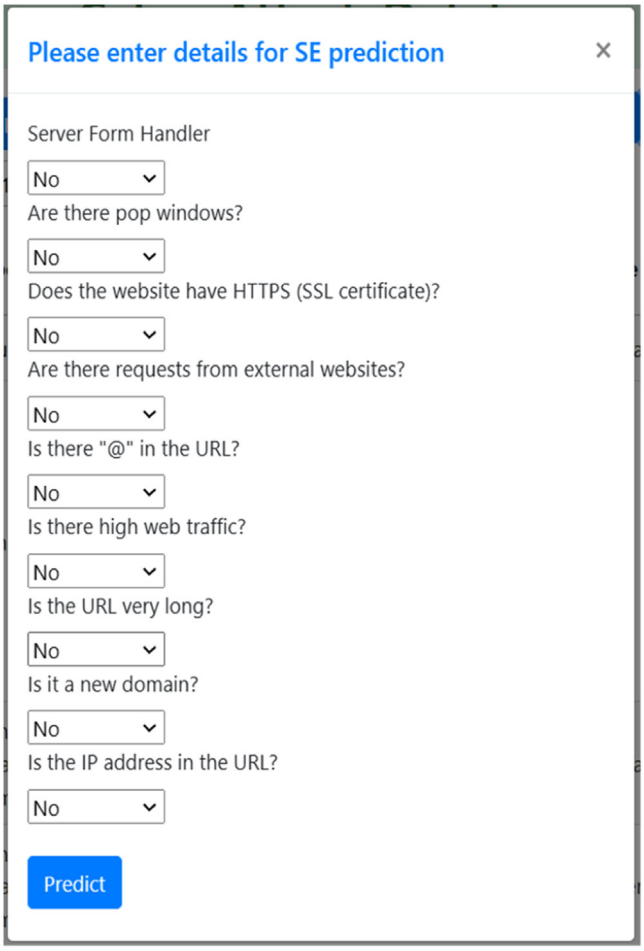

Fig. 10. Graphical user interface.

of a desired result. Depending on the parameters entered by the user, a result of legitimate, suspicious, or phishing will be returned.

For the future scope of this work, it is worth mentioning that phishing attacks are very dynamic in nature, and as soon as cyber criminals are contained in their traps, they exploit other methods to maneuver the problem in order to gain an advantage. This work greatly relies on known intrinsic attributes found on websites and selects key features that are prevalent with substantial correlation with the legitimate or illegitimate status of unknown websites passed through the ML algorithm. Given the explosion of web technologies such as Django, Meteor JS, Yii, and Motion UI, these nine selected features may not be sufficient to accommodate the newer trends in websites used by cyber criminals. Therefore, the scope of this work for the future will be to explore newer trends in website phishing and continually make provision to evolve our web application and database to capture such peculiarities.

6. Limitation

The scope of this work primarily focuses on website phishing, even though there are other forms of phishing attacks such as spearphishing and email phishing that can potentially be devastating. The SVM prediction model is based on a dataset of about 1400 records, which may have had an impact on the model's accuracy; however, this work has remediated that by providing a function that will allow that dataset to grow through a database and that will enable the machine learning model to consistently correct itself, which will in turn improve the accuracy of the model.

7. Conclusion

Phishing attacks continue to pose a significant threat to website security, but there are effective methods for detecting and preventing them. Since it is a social engineering attack that exploits the naivety of a user, building a web application that is user-driven, as this paper suggests, can help curb the menace that this kind of attack persistently poses. There has been a large body of work proposing different methods; however, the phishing trend still remains endemic. Not only has this work made an effort to understand website phishing attacks on a deeper level, it has also stimulated the awareness of users from a security perspective to understand the features of a typical website that they should be suspicious of.

Declaration of competing interest

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

CRediT authorship contribution statement

Gilles Dusserre: Supervision, Writing – review & editing. **Robert Bestak:** Supervision, Writing – review & editing. **Nasir Baba Ahmed:** Investigation, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

References

[1] A.K. Jain, N. Debnath, A.K. Jain, APuML: an efficient approach to detect mobile phishing webpages using machine learning, Wirel. Pers. Commun. 125 (4) (2022) 3227–3248 Aug, doi:10.1007/s11277-022-09707-w.
[2] A. Yasin, R. Fatima, L. Liu, J. Wanga, R. Ali, Z. Wei, Counteracting social engineering attacks, Comput. Fraud Secur. 2021 (10) (2021) 15–19 Oct, doi:10.1016/S1361-3723(21)00108-1.
[3] A. Mughaid, S. AlZu'bi, A. Hnaif, S. Taamneh, A. Alnajjar, E.A. Elsoud, An intelligent cyber security phishing detection system using deep learning techniques, Clust. Comput. 25 (6) (2022) 3819–3828 Dec, doi:10.1007/s10586-022-03604-4.
[4] P. Suresh, et al., Chapter 10 - contemporary survey on effectiveness of machine and deep learning techniques for cyber security, in: Machine Learning for Biometrics, Academic Press, 2022, pp. 177–200, doi:10.1016/B978-0-323-85209-8.00007-9. P. P. Sarangi, M. Panda, S. Mishra, B. S. P. Mishra, and B. Majhiin Cognitive Data Science in Sustainable Computing.
[5] K.C. Bourne, Chapter 15 - security, in: Application Administrators Handbook, Morgan Kaufmann, Boston, 2014, pp. 242–267, doi:10.1016/B978-0-12-398545-3.00015-7. K. C. Bourne.
[6] S.D. Applegate, Social engineering: hacking the wetware!, Inf. Secur. J. Glob. Perspect. 18 (1) (2009) 40–46 Feb, doi:10.1080/19393550802623214.
[7] K. Chetioui, B. Bah, A.O. Alami, A. Bahnasse, Overview of social engineering attacks on social networks, Procedia Comput. Sci. 198 (2022) 656–661 Jan, doi:10.1016/j.procs.2021.12.302.
[8] A. van der Merwe, M. Loock, M. Dabrowski, Characteristics and responsibilities involved in a phishing attack, in: Proceedings of the 4th International Symposium on Information and Communication Technologies, in WISICT '05, Cape Town, South Africa, Trinity College Dublin, 2005, pp. 249–254. Jan.
[9] Z. Alkhalil, C. Hewage, L. Nawaf, I. Khan, Phishing attacks: a recent comprehensive study and a new anatomy, Front. Comput. Sci. 3 (2021) Accessed: Nov. 13, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcomp.2021.563060.
[10] M. Sánchez-Paniagua, E. Fidalgo, E. Alegre, R. Alaiz-Rodríguez, Phishing websites detection using a novel multipurpose dataset and web technologies features, Expert Syst. Appl. 207 (2022) 118010 Nov, doi:10.1016/j.eswa.2022.118010.
[11] J. Rushton, "50+ phishing statistics you need to know – where, Who & What is Targeted," Techopedia. Accessed: Nov. 13, 2023. [Online]. Available: https://www.techopedia.com/phishing-statistics.
[12] R.M. Mohammad, F. Thabtah, L. McCluskey, Tutorial and critical analysis of phishing websites methods, Comput. Sci. Rev. 17 (2015) 1–24 Aug, doi:10.1016/j.cosrev.2015.04.001.
[13] K.L. Chiew, K.S.C. Yong, C.L. Tan, A survey of phishing attacks: their types, vectors and technical approaches, Expert Syst. Appl. 106 (2018) 1–20 Sep, doi:10.1016/j.eswa.2018.03.050.
[14] B.B. Gupta, N.A.G. Arachchilage, K.E. Psannis, Defending against phishing attacks: taxonomy of methods, current issues and future directions, Telecommun. Syst. 67 (2) (2018) 247–267 Feb, doi:10.1007/s11235-017-0334-z.
[15] M.M. Alani, H. Tawfik, PhishNot: a cloud-based machine-learning approach to phishing URL detection, Comput. Netw. 218 (2022) 109407 Dec, doi:10.1016/j.comnet.2022.109407.

[16] J. Hong, The state of phishing attacks, Commun. ACM 55 (1) (2012) 74–81 Jan, doi:10.1145/2063176.2063197.

[17] Q. Cui, G.V. Jourdan, G.v. Bochmann, I.V. Onut, SemanticPhish: a semantic-based scanning system for early detection of phishing attacks, in: Proceedings of the 2020 APWG Symposium on Electronic Crime Research (eCrime), 2020, pp. 1–12, doi:10.1109/eCrime51433.2020.9493252. Nov.

[18] "Number of global phishing sites 2022," Statista. Accessed: May 03, 2023. [Online]. Available: https://www.statista.com/statistics/266155/number-of-phishing-domain-names-worldwide/.

[19] G. Apruzzese, et al., The role of machine learning in cybersecurity, Digit. Threats Res. Pract. 4 (1) (2023) 8:1-8:38Mar, doi:10.1145/3545574.

[20] A. Parisi, Hands-on Artificial Intelligence for Cybersecurity: Implement Smart AI Systems for Preventing Cyber Attacks and Detecting Threats and Network Anomalies, Packt Publishing, Birmingham, UK, 2019.

[21] H. Karimipour, F. Derakhshan, AI-Enabled Threat Detection and Security Analysis for Industrial IoT, Springer Nature, 2021.

[22] Information and Operations Management Department Texas A&M University, R. Sen, G. Heim, Information and Operations Management Department Texas A&M University, Q. Zhu, and Information and Operations Management Department Texas A&M University, Artificial intelligence and machine learning in cybersecurity: applications, challenges, and opportunities for MIS academics, Commun. Assoc. Inf. Syst. 51 (1) (2022) 179–209, doi:10.17705/1CAIS.05109.

[23] M. Wazid, A.K. Das, V. Chamola, Y. Park, Uniting cyber security and machine learning: advantages, challenges and future research, ICT Express 8 (3) (2022) 313–321 Sep, doi:10.1016/j.icte.2022.04.007.

[24] R. Kaur, D. Gabrijelčič, T. Klobučar, Artificial intelligence for cybersecurity: literature review and future research directions, Inf. Fusion 97 (2023) 101804 Sep, doi:10.1016/j.inffus.2023.101804.

[25] A. Safi, S. Singh, A systematic literature review on phishing website detection techniques, J. King Saud Univ. Comput. Inf. Sci. 35 (2) (2023) 590–611 Feb, doi:10.1016/j.jksuci.2023.01.004.

[26] I. Qabajeh, F. Thabtah, F. Chiclana, A recent review of conventional vs. automated cybersecurity anti-phishing techniques, Comput. Sci. Rev. 29 (2018) 44–55 Aug, doi:10.1016/j.cosrev.2018.05.003.

[27] A. Basit, M. Zafar, X. Liu, A.R. Javed, Z. Jalil, K. Kifayat, A comprehensive survey of AI-enabled phishing attacks detection techniques, Telecommun. Syst. 76 (1) (2021) 139–154 Jan, doi:10.1007/s11235-020-00733-2.

[28] G. Vrbančič, I. Fister jr, and V. Podgorelec, "Swarm intelligence approaches for parameter setting of deep learning neural network: case study on phishing websites classification," Jun. 2018, pp. 1–8.

[29] J. James, S. L., and C. Thomas, "Detection of phishing URLs using machine learning techniques," Dec. 2013, pp. 304–309.

[30] S.W. Liew, N.F.M. Sani, Mohd.T. Abdullah, R. Yaakob, M.Y. Sharum, An effective security alert mechanism for real-time phishing tweet detection on Twitter, Comput. Secur. 83 (2019) 201–207 Jun, doi:10.1016/j.cose.2019.02.004.

[31] A. Begum, S. Badugu, A study of malicious URL detection using machine learning and heuristic approaches, Learn. Anal. Intell. Syst. (2019) 587.

[32] R. Fatima, A. Yasin, L. Liu, J. Wang, How persuasive is a phishing email? A phishing game for phishing awareness, J. Comput. Secur. 27 (6) (2019) 581–612 Jan, doi:10.3233/JCS-181253.

[33] A. Pandey, N. Gill, K. Sai Prasad Nadendla, I.S. Thaseen, Identification of phishing attack in websites using random forest-SVM hybrid model, in: Intelligent Systems Design and Applications, Springer International Publishing, Cham, 2020, pp. 120–128, doi:10.1007/978-3-030-16660-1_12. vol. 941A. Abraham, A. K. Cherukuri, P. Melin, and N. GandhiAdvances in Intelligent Systems and Computing, vol. 941.

[34] M. Cova, C. Kruegel, G. Vigna, There is no free phish: an analysis of 'free' and live phishing kits, in: Proceedings of the 2nd Conference on USENIX Workshop on Offensive Technologies, in WOOT'08, USA, USENIX Association, 2008, pp. 1–8. Jul.

[35] T. Moore, R. Clayton, Examining the impact of website take-down on phishing, in: Proceedings of the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit, in eCrime '07, New York, NY, USA, Association for Computing Machinery, 2007, pp. 1–13, doi:10.1145/1299015.1299016. Oct.

[36] S. Minocha, B. Singh, A novel phishing detection system using binary modified equilibrium optimizer for feature selection, Comput. Electr. Eng. 98 (2022) 107689 Mar, doi:10.1016/j.compeleceng.2022.107689.

[37] C.L. Tan, K.L. Chiew, K. Wong, S.N. Sze, PhishWHO: Phishing webpage detection via identity keywords extraction and target domain name finder, Decis. Support Syst. 88 (2016) 18–27, doi:10.1016/j.dss.2016.05.005.

[38] N. Abdelhamid, A. Ayesh, F. Thabtah, Phishing detection based associative classification data mining, Expert Syst. Appl. 41 (13) (2014) 5948–5959 Oct, doi:10.1016/j.eswa.2014.03.019.

[39] M.A. Jabbar, B.L. Deekshatulu, P. Chandra, Knowledge discovery using associative classification for heart disease prediction, Adv. Intell. Syst. Comput. 182 (2013) 29–39 AISC, doi:10.1007/978-3-642-32063-7_4.

[40] F. Thabtah, P. Cowling, Y. Peng, MCAR: multi-class classification based on association rule, in: Proceedings of the 3rd ACS/IEEE International Conference on Computer Systems and Applications, 2005, 2005, p. 33, doi:10.1109/AICCSA.2005.1387030. Jan.

[41] G. Costa, R. Ortale, E. Ritacco, X-Class: Associative classification of XML documents by structure, ACM Trans. Inf. Syst. 31 (1) (2013), doi:10.1145/2414782.2414785.

[42] L.A.T. Nguyen, B.L. To, H.K. Nguyen, M.H. Nguyen, Detecting phishing web sites: a heuristic URL-based approach, in: Proceedings of the 2013 International Conference on Advanced Technologies for Communications ATC 2013, 2013, pp. 597–602, doi:10.1109/ATC.2013.6698185. Oct.

[43] Y. Zhang, J.I. Hong, L.F. Cranor, Cantina: a content-based approach to detecting phishing web sites, in: Proceedings of the 16th International Conference on World Wide Web, in WWW '07, New York, NY, USA, Association for Computing Machinery, 2007, pp. 639–648, doi:10.1145/1242572.1242659. May.

[44] A.K. Jha, R. Muthalagu, P.M. Pawar, Intelligent phishing website detection using machine learning, Multimed. Tools Appl. (2023) Feb, doi:10.1007/s11042-023-14731-4.

[45] A.K. Jain, S. Parashar, P. Katare, I. Sharma, PhishSKaPe: a content based approach to escape phishing attacks, Procedia Comput. Sci. 171 (2020) 1102–1109 Jan, doi:10.1016/j.procs.2020.04.118.

[46] B. Wardman, T. Stallings, G. Warner, A. Skjellum, High-performance content-based phishing attack detection, in: 2011 eCrime Researchers Summit, IEEE, San Diego, USA, 2011, pp. 1–9, doi:10.1109/eCrime.2011.6151977. Nov.

[47] K. Komiyama, T. Seko, Y. Ichinose, K. Kato, K. Kawano, H. Yoshiura, In-depth evaluation of content-based phishing detection to clarify its strengths and limitations, in: U- and E-Service, Science and Technology, Springer, Berlin, Heidelberg, 2010, pp. 95–106, doi:10.1007/978-3-642-17644-9_11. T. Kim, J. Ma, W. Fang, B. Park, B.H. Kang, and D. Ślęzakin Communications in Computer and Information Science.

[48] S. Afroz, R. Greenstadt, PhishZoo: detecting phishing websites by looking at them, in: Proceedings of the 2011 IEEE 5th International Conference on Semantic Computing, Palo Alto, CA, USA, IEEE, 2011, pp. 368–375, doi:10.1109/ICSC.2011.52. Sep.

[49] A. Abuzuraiq, M. Alkasassbeh, M. Almseidin, Intelligent methods for accurately detecting phishing websites, in: Proceedings of the 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, IEEE, 2020, pp. 085–090, doi:10.1109/ICICS49469.2020.239509. Apr.

[50] P.M. Al-kasassbeh, Intelligent methods for accurately detecting phishing websites, in: Proceedings of the 2020 11th International Conference on Information and Communication Systems ICICS, 2020 Jan.Accessed: Nov. 13, 2023. [Online]. Available:.

[51] K.L. Chiew, E. Chang, S. Sze, W. Tiong, Available online utilisation of website logo for phishing detection, Comput. Secur. 54 (2015) Aug, doi:10.1016/j.cose.2015.07.006.

[52] H.Y.A. Abutair, A. Belghith, Using case-based reasoning for phishing detection, Procedia Comput. Sci. 109 (2017) 281–288 Jan, doi:10.1016/j.procs.2017.05.352.

[53] D. Sahoo, C. Liu, and S. C. H. Hoi, "Malicious URL detection using machine learning: a survey." arXiv, Aug. 21, 2019.

[54] Z.H. Zhou, Ensemble learning, in: Machine Learning, Springer, Singapore, 2021, pp. 181–210, doi:10.1007/978-981-15-1967-3_8. Z.H. Zhou.

[55] M. Al-Sarem, et al., An optimized stacking ensemble model for phishing websites detection, Electronics 10 (11) (2021) 11 Art. no.Jan, doi:10.3390/electronics10111285.

[56] J. Abawajy, A. Kelarev, A multi-tier ensemble construction of classifiers for phishing email detection and filtering, in: Cyberspace Safety and Security, Springer, Berlin, Heidelberg, 2012, pp. 48–56, doi:10.1007/978-3-642-35362-8_5. Y. Xiang, J. Lopez, C.C. J. Kuo, and W. Zhouin Lecture Notes in Computer Science.

[57] P. Bountakas, C. Xenakis, HELPHED: hybrid ensemble learning phishing email detection, J. Netw. Comput. Appl. 210 (2023) 103545 Jan, doi:10.1016/j.jnca.2022.103545.

[58] "Choosing the right estimator," scikit-learn. Accessed: Nov. 14, 2023. [Online]. Available: https://scikit-learn/stable/tutorial/machine_learning_map/index.html.

[59] "1.1. Linear Models," scikit-learn. Accessed: Nov. 14, 2023. [Online]. Available: https://scikit-learn/stable/modules/linear_model.html.

[60] D. Anguita, A. Ghio, N. Greco, L. Oneto, S. Ridella, Model selection for support vector machines: advantages and disadvantages of the machine learning theory, in: Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN), 2010, pp. 1–8, doi:10.1109/IJCNN.2010.5596450. Jul.

[61] "4. Supervised learning: models and concepts - machine learning and data science blueprints for finance [Book]." Accessed: Nov. 14, 2023. [Online]. Available: https://www.oreilly.com/library/view/machine-learning-and/9781492073048/ch04.html.

[62] E.A. Zanaty, Support vector machines (SVMs) versus multilayer perception (MLP) in data classification, Egypt. Inform. J. 13 (3) (2012) 177–183 Nov, doi:10.1016/j.eij.2012.08.002.

[63] C.W. Hsu, C.J. Lin, A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Netw. 13 (2) (2002) 415–425, doi:10.1109/72.991427.

[64] N. Abdelhamid, "Website phishing." UCI Machine Learning Repository, 2014.

[65] I.H. Witten, E. Frank, Data mining: practical machine learning tools and techniques with Java implementations, ACM SIGMOD Rec. 31 (1) (2002) 76–77 Mar, doi:10.1145/507338.507355.

[66] R.M. Mohammad, F. Thabtah, L. McCluskey, An assessment of features related to phishing websites using an automated technique, in: Proceedings of the 2012 International Conference for Internet Technology and Secured Transactions, 2012, pp. 492–497. Dec.

[67] S.N. Wan Ahmad, Comparative performance of machine learning methods for classification on phishing attack detection, Int. J. Adv. Trends Comput. Sci. Eng. 9 (1.5) (2020) 349–354 Sep, doi:10.30534/ijatcse/2020/4991.52020.

[68] D. K. Srivastava and L. Bhambhu, "Data classification using support vector machine," 2005.

[69] D. Wahyudi, M. Niswar, A.A.P. Alimuddin, Website phising detection application using support vector machine (SVM), J. Inf. Technol. Its Util. 5 (1) (2022) 18–24 Jun, doi:10.56873/jitu.5.1.4836.

[70] M. Nabet, L. George, Phishing attacks detection by using support vector machine, J. Al-Qadisiyah Comput. Sci. Math. 15 (2023) Sep, doi:10.29304/jqcm.2023.15.2.1242.

[71] D. Aksu, A. Abdulwakil, and M. A. Aydin, "Detecting phishing websites using support vector machine algorithm," presented at the Pressacademia, Jun. 2017, pp. 139–142. doi:10.17261/Pressacademia.2017.582.

[72] A. Altaher, Phishing websites classification using hybrid SVM and KNN approach, Int. J. Adv. Comput. Sci. Appl. 8 (6) (2017), doi:10.14569/IJACSA.2017.080611.

[73] R. Karnik and D. G. M. Bhandari, “Support vector machine based malware and phishing website detection,” 2016. Accessed: Nov. 14, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Support-Vector-Machine-Based-Malware-and-Phishing-Karnik-Bhandari/ffea603ec9f33931c9de630ba1a6ac71924f1539.

[74] A. Mandadi, S. Boppana, V. Ravella, R. Kavitha, Phishing website detection using machine learning, in: Proceedings of the 2022 IEEE 7th International Conference for Convergence in Technology (I2CT), 2022, pp. 1–4, doi:10.1109/I2CT54291.2022.9824801. Apr.

[75] A.K. Dutta, Detecting phishing websites using machine learning technique, PLoS ONE 16 (10) (2021) e0258361 Oct, doi:10.1371/journal.pone.0258361.

[76] S. Alnemari, M. Alshammari, Detecting phishing domains using machine learning, Appl. Sci. 13 (8) (2023) 8 Art. no.Jan, doi:10.3390/app13084649.

[77] Z. Alshingiti, R. Alaqel, J. Al-Muhtadi, Q.E.U. Haq, K. Saleem, M.H. Faheem, A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN, Electronics 12 (1) (2023) 1 Art. no.Jan, doi:10.3390/electronics12010232.

[78] Md.A.A. Siddiq, M. Arifuzzaman, M.S. Islam, Phishing website detection using deep learning, in: Proceedings of the 2nd International Conference on Computing Advancements, in ICCA ’22, New York, NY, USA, Association for Computing Machinery, 2022, pp. 83–88, doi:10.1145/3542954.3542967. Aug.