# The Survey and Proposal on Machine Learning Based Phishing Detection Techniques

**Lokendra Singh Songare[1], Dr. Dhanraj Verma[2]**

Research Scholar[1], Professor[2]

Department of Computer Science Engineering, Dr. APJ Abdul Kalam University, Indore (M.P.)

lokendra.songare@gmail.com

dhanrajmtech@gmail.com

**Abstract:** Rapidly growing cyber infrastructure has reduced the cost of communication and internettherefore,normal people and attackers usages it much frequently. Additionally cheaters are changing their faces therefore secure communication is essential.Among a number of security challenges phishing is one of the issues where a fixed solution is not available. Therefore identification of changing faces of phishing is needed to be investigated. In this work the current techniques based on ML (machine learning)are investigated and categorize methods according to the features used and classifiers. Using the observed techniques a model based ML is proposed. The basic design of the required phishing detection model is reported. Additionally its functional aspects are discussed. Finally the summary is provided and future work is proposed.

**Keywords:cyber security, email, fraud, phishing, URL classification, emails classification.**

## I. INTRODUCTION

Phishing is one of the popular techniques used by attackers with the intention of exploiting the internet user'spersonal details. It is a form of identity theft attack occurs when a malicious web site impersonate a user to get sensitive or private information i.e. passwords, account details, or others. There aresome anti-phishing techniques for detecting phishing attempts by evaluation of emails and contents on websites. Phishers come up with new techniques. It is a social engineering that bypassessecurity implemented to mitigate risks. It capitalizes this weakness and exploits human nature to gain access. At the beginning of phishing, attackers were acting alone [1]. As financial organizations have increased their on-line presence, the value of on-line account has increased. These attacks have become more and more professional, organized and systematic. The focus of attacks have turned to consumers of online banks, retailers and service providers. The media of phishing is usually online i.e. e-banks, Internet Relay Chatting (IRC), Instant Messaging (IM), and Email.

Mostly, attacker poses as an employee of an organization, gains trust from the consumers, and then consumers send out their sensitive dataor attacker create fake websites to increase the phishing. For exampleattacker registersa number of domains that are similar to popular brand, i.e. *"www.cit1bank.com"* or *"www.citi-bank.com"*. Victims, enter one of these websites, and believe that the website is real, and try to operate their account [2].Therefore, security is essential to prevent the data from attacks. It may active or passive attack. The passive phishingattack is a threat and larger in social media too, i.e. Facebook and Twitter. Phishing using emails contain link to the infected website where user asked to enter the personal information, so the information can be used by attacker. The email is send to large number of people and the attacker will count the percentage of people who read email and entered data. It is difficult to find that we are visiting an actual site or malicious [3].

In this presented work the main aim is to enhance the existing phishing detection techniques andto offer full proof security against the phishing attacks. In this context the following objectives are established:

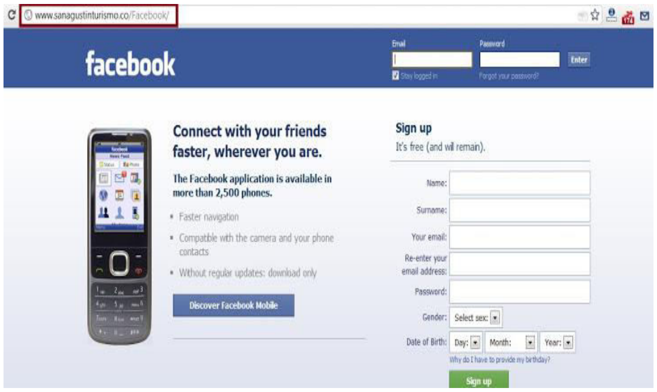
1. **Exploring and investigating the effective and less resource consuming techniques for anti-phishing tool design:** To design an improved technique for phishing detection, need to understand the nature of phishing and deployment technique. Therefore categorization and evaluation of existing methods and techniques is essential. This phase involve the theoretical investigation of the existing anti-phishing tool design.
2. **Designing and developing the enhanced anti-phishing tool:** using the collected literature and the observed techniques is concluded to design and implement an appropriate tool. That technique helps to improve accuracy of existing models by new features and strategies. Additionally it includes the comparative study with respect to the available techniques.

3. **Extending the technique of proposed phishing detection:** Using the extended featuresfrom URLs, the previously proposed technique is extended for improving the performance and reducing resource consumption. The new technique also extended here for adopting the new patterns of phishing deployments.
4. **Comparative performance study tojustifying the proposed phishing detection model:** the proposed technique for phishing identification is compared against the existing techniques for justifying the proposed work outcomes. In addition of that by using the observations and experience of the system design the advantages and drawbacks of the proposed work is delivered.

This section provides the overview of the proposed work involved. The next section provides the basic concepts relevant to proposed study.

## II. BACKGROUND

For example, a page which looks like Facebook but it has a different URL. A user lands on this page, might think it is real page and can provide their id and password. So who don't find the login page is suspicious might enter their id, password and would be sent to the hacker [4]. It is affecting major sectors of industry with a lot of misuse. The phishing attackers trick users by different social engineering tactics such as to suspend accounts, account update process, other information to validate the accounts or some other reasons to get the users to visit spoofed pages [5].



www.sanagustinturismo.co/Facebook/

Connect with your friends faster, wherever you are.

The Facebook application is available in more than 2,500 phones.

- Faster navigation
- Compatible with the camera and your phone contacts
- Without regular updates: download only

Sign up

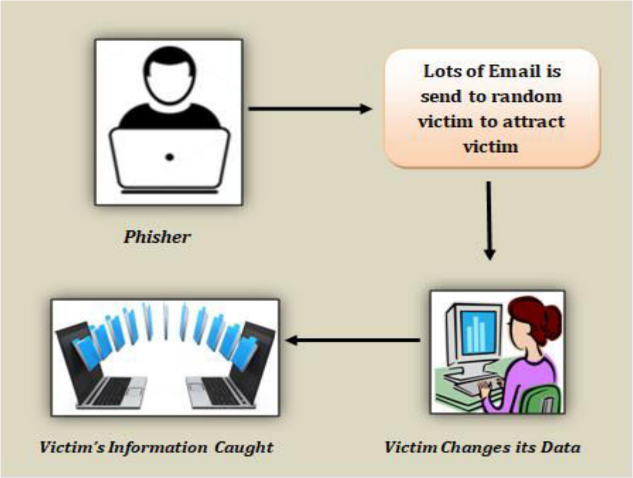
It's free (and wil remain).

Name:
Surname:
Your email:
Re-enter your email address:
Password:
Gender: Select sex:
Date of Birth: Day: Month: Year:

Why do I have to provide my birthday?

Figure 1 Example of Facebook Phishing

One of the primary goals of phishing is to carry out fraudulent financial transactions on behalf of users that contain a URL pointing to a fake web site.A phisher may lure a victim into giving his/her Social Security Number, full name, and address. That can be used for a credit card on the victim's behalf [6]. Phishing attacks are deployedin following steps:



Lots of Email is send to random victim to attract victim
Phisher
Victim's Information Caught
Victim Changes its Data

Figure 2 Process of Phishing

- A fake web site looks like the original Web site is set up.

- Then link sent of the fake web site using e-mails or other messaging system to target users by the name of a brand, trying to convince the victims to visit their web page.
- Victims visit the fake web page by clicking on the link and input their information.
- Then steal the information to do fraud such as transferring money from victims' account.

### Types of Phishing Attacks

Here, different types of phishing attack available [7] [8] [9].

A. **Deceptive Phishing:** it is a messages that is confirm information about the account, requesting users to reenter their information, fictitious account charges, unwanted account changes, new free services requiring, quick action, and many other malicious sites are send to many recipients with the hope that the unsuspecting will react by clicking a link or signing onto a fake site.

B. **Malware-Based Phishing:** This scam involve malicious software on users' PCs. Malware can be as an email attachment, or a downloadable file that are not always able to keep their software up to date.

C. **Key loggers and Screen loggers:**This type of malware tracks the input from the keyboard and sends to the hackers. They go into the users' browsers as a program and run automatically with browser.

D. **Session Hijacking:** This monitoring the activities of the users until they sign in to the account or transaction. At that time, the infected software will perform unauthorized actions, like transferring funds, without the user's knowledge.

E. **Data Theft:** Sensitive data’s will be stored in Pcs. That data’s will be taken by the attacker without knowing to the user. This information is user credentials i.e. passwords, social security numbers, credit card, and other, or other confidential information by stealing communications, documents, legal opinions, etc., thieves profit from selling to those who may want to embarrass or cause economic damage.

F. **DNS-Based Phishing:** DNS based phishing is modify the hosts file. The hackers will return a bogus address and the communication will be sent to fake website. Users are unaware of this and will enter the personal information and it will be hacked by the hackers.

G. **Search Engine Phishing:**Phishers create sites for fake products, they get the pages indexed by search engines, then await unsuspecting customers to enter tip as a part of an order, sign-up, or balance transfer

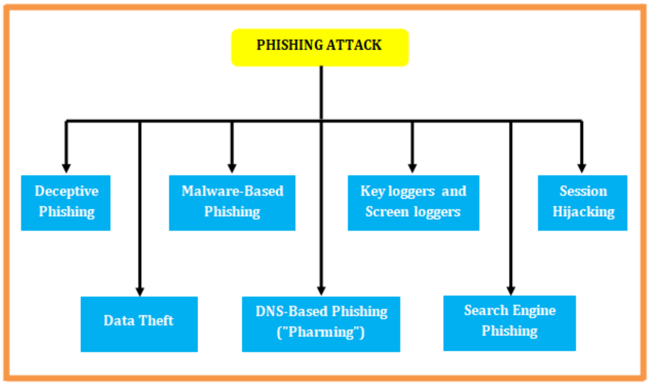


Figure 3 Phishing Attack Classifications

## III. LITERATURE SURVEY

This section provides understanding about the recent development and investigation on the phishing detection and URL classification models.

***R. Verma et al [10]*** building a system for URL analysis and classification to detect phishing attacks. URL analysis is maintaining distance between the attacker and the victim, rather than visiting the website. They research a few realities of URL examination, e.g., execution investigation on both adjusted and lopsided datasets, in a static and live exploratory arrangement and online versus clump learning.***P. Patil et al [11]*** aim to use visual features of a web page’s as the basis of detecting page similarities. They propose a solution, to detect phishing web pages. Page layouts and contents are the feature of web pages’. To specify page layout the style sheet (CSS) is used, they develop an algorithm to detect similarities in key elements into CSS. And proposed a system using SVM with map-reduce to achieve a higher accuracy in spam email detection.***M. Moghimi et al [12]*** present a rule-based method to detect phishing in internet banking. That used two features, which have been determining the webpage identity. The features include four features to evaluate the page, and four features to access protocol. They are using string matching algorithms to determine the relationship between content and URL. They employed SVM to classify webpages. The experiments indicate the model can detect phishing with accuracy of 99.14%. Sensitivity analysis demonstrates the impact of features.

Phishing is one of the most severe cybercrimes. Butthere is no complete and accurate solution. ***G. Varshney et al [13]*** studies, analyzes, and classifies significant strategies for phished website detection, and outline advantages and drawbacks. Furthermore, an analysis of the latest schemes in various categories is provided. They identify advantages, drawbacks, and research gaps. The analysis will help academia and industries.***S. Gupta et al [14]***, a dynamic approach based on minimum time to detect Phishing URL by using classification. This approach is based on different parameters i.e. accuracy, Recall, Precision, Specification, and many more. The result shows that Random tree is a good classification technique.***B. Wei et al [15]*** designs an accurate and low-cost phishing detection sensor by deep learning. They proposed a light-weight deep learning algorithm to detect the malicious URLs and enable a real-time phishing detection sensor. Experimental tests and comparisons have been verifying the efficiency of the method. According to that true detection rate has been improved. They also verified the method can run in an energy-saving single board computer.

The accuracy of detection depends on prior knowledge. Features extracted from different dimensions are comprehensive and time consuming. ***P. Yang et al [16]*** propose a multidimensional feature phishing detection approach based on deep learning. First, character sequence of URL is used for classification, and not requires third party assistance. Second, combine URL features, code features, text features, and classification result. That can reduce the detection time. Testing on phishing and legitimate URLs, the accuracy reaches 98.99%.***G. Sonowal et al [17]*** provides a multilayer model to detect phishing, titled as PhiDMA. That incorporates five layers: Auto upgrade white-list layer, URL features layer, Lexical signature layer, String matching layer and Accessibility Score comparison layer. A prototype implementation of the PhiDMA is built with an accessible interface. The result shows that the model is capable to detect phishing with an accuracy of 92.72%***A. C. Bahnsen et al [18]*** explored the use of URLs as input for ML models for phishing site prediction. They compared a feature-engineering approach by a random forest classifier against a method based on RNN. They determined RNN provides an accuracy of 98.7% without the need of feature.

Visual similarities based techniques are useful for detecting phishing. That techniques utilize the feature set like text content, text format, HTML tags, CSS, image, and so forth. These approaches compare the suspicious website with the legitimate website using features and similarity greater than a threshold is declared phishing. ***A. K. Jain et al [19]*** presents an analysis of attacks, their exploitation, recent visual similarity based approaches, and comparative study. There are two concerns with existing approaches. First the large number of training features and the lack of arguments. Second the type of datasets that are biased with features. ***H. Shirazi et al [20]*** put forward sign of phishing and holds the key to successful detection. They design features that model the relationships, visuals and statistics. The value of feature design is to makehard to tamper. The model trains with seven features and achieves a true positive rate of 98% and a classification accuracy of 97%. Data classification is 4 times faster for legitimate and 10 times faster for phishing.Many techniques use source code-based features and third party services. These techniques have some limitations. The third-party services delays the classification. ***R. S. Rao et al [21]*** propose a light-weight, CatchPhishto predict the URL legitimacy. That uses hostname, URL, TF-IDF and phish hinted words for Random forest. The model with TF-IDF achieved accuracy of 93.25% and with TF-IDF and hand-crafted features achieved 94.26%.

***A. E. Aassal et al [22]*** perform a study and evaluate phishing features on diverse datasets and propose a new taxonomy of features. Next, propose a structure called 'PhishBench,' which empowers us to assess and think about the highlights, i.e., framework detail, datasets, classifiers, and assessment measurements.That is a first benchmarking phishing related research and evaluation for feature comparison. They use it to test methods on new datasets. They study how dataset characteristics, e.g., varying legitimate to phishing ratios and increasing size of imbalanced datasets, affect classification performance. The results show the imbalanced attacks affect the detection and retraining alone is not enough. Most of the approaches are feature based and cannot detect dynamic attacks. The attacker uses the input form, content and embeds @ symbol in URL. To detect this, Behaviour based Malicious URL Finder is proposed by ***N. Jayakanthan et al [23]***. It analyzes the behaviour of the URL. The FSM based state transition is used to model the URL behaviour. The state transition from initial to final state is used for classification. This approach tests the genuine and malicious behavior of the URL.Phishing detection algorithms can be an effective approach to safeguarding users from such attacks. ***A. M. A. Zuraiq et al [24]*** will review different phishing detection approaches which include: Content-Based, Heuristic-Based, and Fuzzy rule-based approaches.

***P. Yi et al [25]*** focuses on applying deep learning to detect phishing websites. First design two types of features: original features and interaction features. A detection model based on Deep Belief Networks is presented andtested using real IP flows from ISP.The detecting model achievesa 90% true positive rate and 0.6% false positive rate.The blacklist-based approach has proven inefficient. Considering hybrid intelligent approach based on rule induction for phishing detection is still an open issue. The algorithm capable of separating phishing websites is proposed by ***K. S. Adewole et al [26]***. The algorithm leverages the strengths of JRip and Projective Adaptive Resonance Theoryto generate rules. Experiments on two datasets demonstrateitachieving accuracy of 0.9453. ***W. Wang et al [27]*** propose a fast phishing detection approach called PDRCNN that relies only URL. It encodes the URL into a 2D tensor and feeds into a deep learning NN to classify. A bidirectional LSTM used to extract global features and convert string to character. and use a CNN to judge which characters roles in phishing detection, capture the key components, and compress the features. They built a dataset using Alexa and PhishTank. Results show that it achieves a detection accuracy of 97%.

***Jr. T. Chin et al [28]*** present PhishLimiter, a detection and mitigation approach.First a deep packet inspection (DPI) and leverage it with software-defined networking (SDN) to identify phishing activities. It consists: phishing signature and real-time DPI. Using SDN, they develop, store and forward mode.The mode used to the direct network traffic using an ANN. It provides better traffic management for phishing attacks. They usedata sets of real email with links. The experiment shows that it provides an effective and efficient solution.***A. Das et al [29]*** reexamine the existing research on phishing.Challenges are: real-time detection, active attacker, dataset quality and base rate fallacy. This consolidates the literature and illuminates opportunities. They organize the literature based on techniques for different attack (e.g., URLs, websites, emails). For detection techniques they examine properties of the dataset, feature extraction, algorithms used, and performance.***A. Kazi et al [30]*** propose a technique to detect and prevent the phishing on e-mail. That uses hyperlink features to detect phishing and use digital signature to prevent. Attack initiated by sending e-mails on the user with links. The application will act as an interface between e-mail and user. This will be more cost effective and better to prevent people.

***M. Kaytan et al [31]*** proposed a model for detecting web pages based on Extreme Learning Machine. And, used a specific web page features to prevent phishing. They have suggested some new rules for features. The model has 30 inputs and 1 output. The validation has been performed. The average classification accuracy was measured as 95.05%.***H. S. Hota et al [32]*** constructing an ensemblemodel to detect phishing E-mail with Remove-Replace Feature Selection Technique (RRFST). It is selecting arandom feature and removes it if accuracy is being unchanged otherwise feature is replaced to original. Classifiers were developed using two algorithms i.e. C4.5 and CART with ensemble with reduced feature. Results indicate that FST produces remarkable performance of 99.27% accuracy using ensemble with only 11 features.The results of deep NN depend on the learning parameters. ***G. Vrbančič et al [33]*** propose a swarm intelligence based approach to parameter setting. By applying to the phishing websites, model was able to improve its detection when compared to existing algorithms.A phishing email is an email which is intended to trick the beneficiary and uncovers touchy data or downloads vindictive programming through connections. ***N. Moradpoor et al [34]*** proposed a NN-based model for detections of phishing emails using publically available email datasets. The results demonstrate the effectiveness of the model in terms of accuracy, true-positive rate, false-positive rate, network performance and error histogram.

## IV. LITERATURE SUMMARY

In order to investigate the phishing design techniques using ML we involve 23 research papers. The contributions of these models are reported in previous section. Here these papers are summarized on the basis of the features type used and type of learning algorithm. the table 1 contains the information about the features and classifier used.

Table 1 summary

| Authors/ publication / year | Features | Classifiers |
|---|---|---|
| R. Verma et al [10], ACM 2017 | URL analysis and classification | Usages distance between attacker and victim. |
| P. Patil et al [11], 2017 IEEE | Visual features for web page (page layouts and contents for layouts CSS used). Spam email detection | SVM with map-reduce to achieve a higher accuracy in spam email detection |
| M. Moghimi et al [12], Expert Systems With Applications, 2016 | String matching algorithms used to determine the relationship of content and URL. | SVM to classify |
| G. Varshney et al [13],Security | Study, analysis, and classifies the | Survey and provide advantages, |

| | | |
|---|---|---|
| Comm. Networks 2016 | significant strategies | drawbacks, and research gaps. |
| S. Gupta et al [14], Soft Computing: Theories and Applications 2016 | Phishing URL classification | Random tree |
| B. Wei et al [15], Sensors 2019 | URLs and real-time detection sensor. | Deep learning |
| P. Yang et al [16], 2019 IEEE | multidimensional features, Character sequence of URL, URL, code, text | Deep learning |
| G. Sonowal et al [17], Computer and Information Sciences, 2020 | A multilayer model with five layers: Auto upgrade white-list layer, URL features layer, Lexical signature layer, String matching layer and Accessibility Score comparison layer. | Deep learning |
| A. C. Bahnsen et al [18], 2017 IEEE | URLs as input | ML models feature-based approach by a random forest classifier against a RNN method. |
| A. K. Jain et al [19], Hindawi 2017 | Visual similarity based feature like text, format, HTML, CSS, image, and other. | threshold based |
| H. Shirazi et al [20], SACMAT' 2018 | domain name of phishing websites is sign of phishing design features that model the relationships, visual and statistical, of the domain name. | Classifier |
| R. S. Rao et al [21], Journal of Ambient Intelligence and Humanized Computing 2020 | uses hostname, URL, TF-IDF and phish hinted words | Random forest classifier. |
| A. E. Aassal et al [22], IEEE 2020 | study and evaluation of features | 'PhishBench,' dataset bench marking |

Based on the obtained outcomes of this review we found that there are various kinds of models, among some of them working on the content of a website for finding the difference between legitimate and phishing site. Additionally some methods are working on URL classification models to prevent the phishing cases. Similarly some of the techniques are based on features based on email contents and links in email. According to our perception the email based phishing prevention is most fit idea to preserve the phishing cases. Therefore in future the email based phishing techniques are used for system design.

Table 2 Features used

| S. No. | Techniques (feature selection) | Used |
|---|---|---|
| 1 | URL based | [10], [12], [14], [15], [17], [18], [20], |
| 2 | Visual features | [11], [19] |

| | | |
|---|---|---|
| 3 | Multi-dimension features | [16], [17], [21] |
| 4 | Survey | [13], [22] |

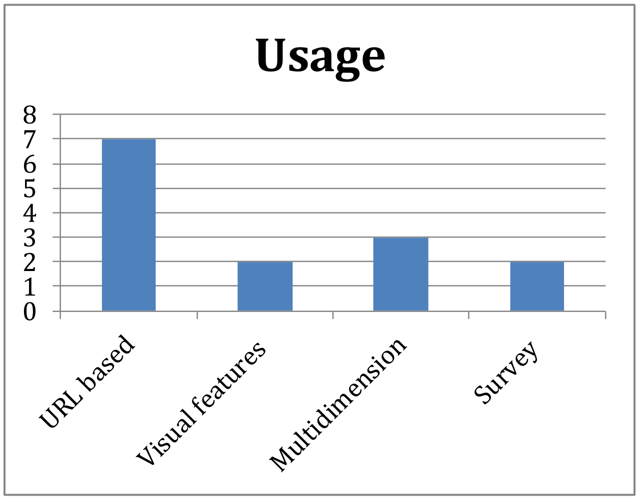


Figure 4 feature selection technique

According to the demonstrated bar graph of features used for phishing identification multi-dimensional features are essential and less explored thus in future for model development the multi-dimensional features are used in our proposed work.Similarly to compute the classes of the URLs, text and other contents for identifying the phishing the classification techniques are used. Thus using the collected literature the frequently used classifiers and learning methods are demonstrated in table 3. This table shows the classification used and their research contribution. Additionally the patterns are reported in figure 5.

Table 3 Classifiers used

| S. No. | Classification method | Used |
|---|---|---|
| 1 | Distance based | [10] |
| 2 | SVM | [11] [12] |
| 3 | Survey | [13], [22] |
| 4 | Random forest | [14], [21] |
| 5 | Deep learning | [16], [17], [18] |
| 6 | Other ML technique | [19], [20] |

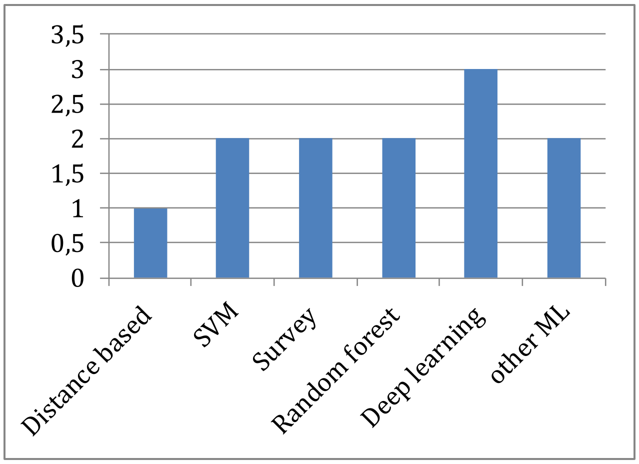


Figure 4 classifiers used

In the survey we observed that the random forest, SVM and Deep Learning based models are much frequently used for developing the phishing detection techniques.

## V. SIMULATION AND RESULTS

In literature we have found that there are much effective technique is to classify the email based phishing attempts. This issue becomes more critical because a number of communication channels are available for text messaging such as WhatsApp, Email, SMS and others. These messages may contain text and links both. Thus the proposed work is focused on text and URL based features classification for phishing identification.Thus based on collected experience a new technique is proposed. A basic proposal of anti-phishing tool is demonstrated using the figure 4. That includes the various functional blocks their overview is reported in this section.

**Learning database:** the ML techniques are required example patterns for learning. These learning samples are built with previously reported real phishing data. In this work we obtain these using online sources i.e. phish tank database and other email dataset. Using these dataset the contents of the source is also located as the training sample. These training samples are also variable and can include more source of information in form of URLs and the web pages.

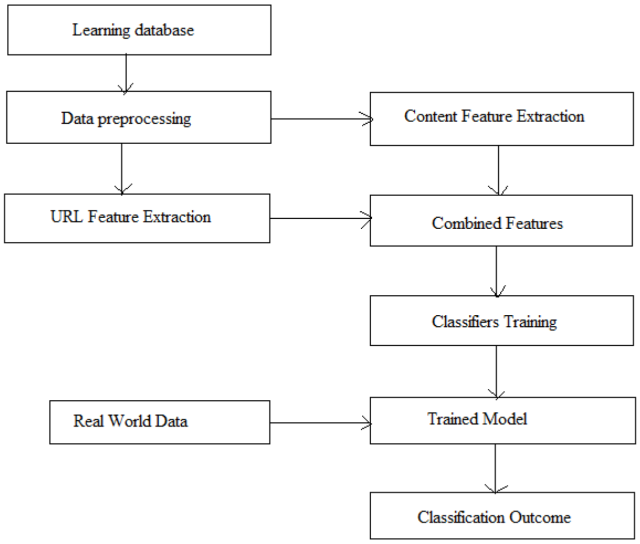


Figure 4 proposed system overview

**Preprocessing:** The preprocessing is an essential step of data mining and ML techniques. These techniques are mostly used for refining the noisy and unwanted contents from the example data. Therefore in this work the URL and web page content are preprocessed for enhancing the data quality and reducing the noise and unwanted content from the data. During preprocessing the URL and text contents are separated in two parts.

**URL feature extraction:** the preprocessing of URL data is performed in this module. Therefore the URL properties are used to read and extract the required features. Therefore the URL is transformed into the 2D vectors and then it is further used for classification task.

**Content feature extraction:** the contents are different from the URL data therefore the features are selected with the help of different approach. Therefore here TF-IDF based features and NLP features are extracted for classification.

**Combined feature computation:** the extracted features using the URLs and the contents are combined. Some of the features are considered as it is, and some of the features are prepared on the basis of the available or extracted features.That are replaced and refined for further investigation.

**Classifier training:** there are various supervised and unsupervised learning techniques are available for classification. However the performance of supervised learning techniques is higher as compared to the unsupervised learning techniques is higher. Therefore first a comparison among recovered classifiers i.e. SVM, random forest and deep NN classification techniques are preformed and accurate and efficient technique is used further.

**Trained model:**The ML techniques are usages the example patterns and prepare a model for finding the similar pattern or behavioral data. The employed classifier trained using the predefined pattern and after learning it is capable to classify the similar patterns in terms of legitimate and phishing URLs.

**Test dataset:** the test dataset is prepared on the basis of existing learning samples as well as some recognized online resources. The test data is applied on the trained model for predicting class label for input emails. Additionally to compute the performance of the anti-phishing tool the test dataset is used. It composed with 30% of randomly selected malicious as well as the legitimate URLs and contents.

**Classification outcomes:** the system is a binary classification technique therefore results legitimate or phishing as outcome. Using these consequences the performance of the anti-phishing technique is provided in terms of memory, time consumption, accuracy and error rate like parameters.

## VI. CONCLUSION & FUTURE WORK

The proposed investigation of phishing detection and classification for emails based attacks.Initially we are selected some essential contributions for review, the review summarized their efforts first and then trends of ML based technique design is investigated. In this context we found the following observations:

1. There are URL based, content based and multi-dimensional feature based techniques mostly used.
2. Most of techniques are using SVM, Random forest and Deep learning based techniques
3. Email based phishing attacks are much serious and effective for preventing them

Therefore to work for future model a basic proposal of the model is also included in this work. That model usages various components which is used for different functional aspects. In order to extend and improve the basic model the following future work is proposed.

1. A comparative study of different classification algorithms for URL classification and also for text classification.
2. Find the different phishing features and their extension techniques
3. Propose a lightweight, efficient and accurate data model for classification
4. Comparative performance study for work justification

## REFERENCES

[1] R. Butler, “Investigation of phishing to develop guidelines to protect the Internet consumer's identity against attacks by phishers”, south African journal Vol.7(3) September 2005

[2] T. N. Jagatic, N. A. Johnson, M. Jakobsson, F. Menczer, “Social Phishing”, Communications of the ACM, October 2007, Vol. 50 No. 10, Pages 94-100 10.1145/1290958.1290968

[3] E. Kirda, C. Kruegel, “Protecting Users Against Phishing Attacks with AntiPhish”, Computer Software and Applications Conference, COMPSAC 2005. 29th Annual International (Volume: 1).

[4] L. Muthiyah, *“What is Phishing? How to Create Phishing Page, Facebook Example”*, available online at: https://www.7xter.com/2016/08/phishing.html. Last Modified: March 20, 2017

[5] R. Basnet, S. Mukkamala, A. H. Sung, *"Detection of phishing attacks: A machine learning approach"*, Soft Computing Applications in Industry, Springer Berlin Heidelberg, PP. 373-383, 2008.

[6] H. Tout, W. Hafner*"Phishpin: An identity-based anti-phishing approach"*, in proceedings of international conference on computational science and engineering, Vancouver, BC, pages 347-352, 2009

[7] I. R. A. HAMID, J. Abawajy, T. Kim, *"Using Feature Selection and Classification Scheme for Automating Phishing Email Detection"*, Studies in Informatics and Control 22(1): pp. 61-70, March 2013

[8] V. Suganya, *"A Review on Phishing Attacks and Various Anti Phishing Techniques"*, International Journal of Computer Applications (IJCA), Volume 139 – No.1, April 2016.

[9] M. Atighetchi, P. Pal, *"Attreibute-based prevention of Phishing Attacks"*, Proceedings of the 8th IEEE International Symposium on Network Computing and Applications, 2009

[10] R. Verma, A. Das, *"What's in a URL: Fast Feature Extraction and Malicious URL Detection"*, IWSPA '17, March 24-24 2017, Scottsdale, AZ, USA. c 2017 Copyright held by the owner/author(s). ACM ISBN 978-1-4503-4909-3/17/03

[11] P. Patil, R. Rane, M. Bhalekar, *"Detecting Spam and Phishing Mails Using SVM and Obfuscation URL Detection Algorithm"*, International Conference on Inventive Systems and Control (ICISC-2017), 978-1-5090-4715-4/17/$31.00 ©2017 IEEE

[12] M. Moghimi, A. Y. Varjani, *"New rule-based phishing detection method"*, Expert Systems With Applications, 53, 2016, 231-242

[13] G. Varshney, M. Misra, P. K. Atrey, *"A survey and classification of web phishing detection schemes"*, Security Comm. Networks 2016; 9:6266–6284

[14] S. Gupta, A. Singhal, *"Dynamic Classification Mining Techniques for Predicting Phishing URL"*, Soft Computing: Theories and Applications, Advances in Intelligent Systems and Computing 584, https://doi.org/10.1007/978-981-10-5699-4_50

[15] B. Wei, R. A. Hamad, L. Yang, X. He, H. Wang, B. Gao, W. L. Woo, *"A Deep-Learning-Driven Light-Weight Phishing Detection Sensor"*, Sensors 2019, 19, 4258; doi:10.3390/s19194258

[16] P. Yang, G. Zhao, P. Zeng, *"Phishing Website Detection Based on Multidimensional Features Driven by Deep Learning"*,Vol. 7, 2169-3536 2019 IEEE

[17] G. Sonowal, K. S. Kuppusamy, *"PhiDMA – A phishing detection model with multi-filter approach"*, Journal of King Saud University – Computer and Information Sciences, 32, 2020, 99-112

[18] A. C. Bahnsen, E. C. Bohorquez, S. Villegas, J. Vargas, F. A. Gonzalez, *"Classifying Phishing URLs Using Recurrent Neural Networks"*, 978-1-5386-2701-3/17/$31.00 c 2017 IEEE

[19] A. K. Jain, B. B. Gupta, *"Phishing Detection: Analysis of Visual Similarity Based Approaches"*, Hindawi Security and Communication Networks Volume 2017, Article ID 5421046, 20 pages

[20] H. Shirazi, B. Bezawada, I. Ray, *""Kn0w Thy Doma1n Name": Unbiased Phishing Detection Using Domain Name Based Features"*, SACMAT'18, June 13-15, 2018, Indianapolis, IN, USA

[21] R. S. Rao, T. Vaishnavi, A. R. Pais, *"CatchPhish: detection of phishing websites by inspecting URLs"*, Journal of Ambient Intelligence and Humanized Computing (2020) 11:813–825

[22] A. E. Aassal, S. Baki, A. Das, R. M. Verma, *"An In-Depth Benchmarking and Evaluation of Phishing Detection Research for Security Needs"*, Special Section on Emerging Approaches To Cyber Security, Vol 8, 2020

[23] N. Jayakanthan, A. V. Ramani, *"Classification Model to Detect Malicious URL via Behaviour Analysis"*, International Journal of Computer Applications Technology and Research Volume 6–Issue 3, 133-140, 2017, ISSN:-2319–8656

[24] A. M. A. Zuraiq, M. Alkasassbeh, *"Review: Phishing Detection Approaches"*, 978-1-7281-2882-5/19/$31.00 ©2019 IEEE

[25] P. Yi, Y. Guan, F. Zou, Y. Yao, W. Wang, T. Zhu, *"Web Phishing Detection Using a Deep Learning Framework"*, Hindawi Wireless Communications and Mobile Computing Volume 2018, Article ID 4678746, 9 pages

[26] K. S. Adewole, A. G. Akintola, S. A. Salihu, N. Faruk, R. G. Jimoh, *"Hybrid Rule-Based Model for Phishing URLs Detection"*, Springer Nature Switzerland AG 2019 All Rights Reserved iCETiC 2019, LNICST 285, pp. 119–135, 2019.

[27] W. Wang, F. Zhang, X. Luo, S. Zhang, *"PDRCNN: Precise Phishing Detection with Recurrent Convolutional Neural Networks"*, Hindawi Security and Communication Networks Volume 2019, Article ID 2595794, 15 pages

[28] Jr. T. Chin, K. Xiong, C. Hu, *"Phishlimiter: A Phishing Detection and Mitigation Approach Using Software-Defined Networking"*, Vol 6, 2018, 2169-3536, 2018 IEEE

[29] A. Das, S. Baki, A. E. Aassal, R. Verma, A. Dunbar, *"SOK: A Comprehensive Reexamination of Phishing Research from the Security Perspective"*, Re-examining Phishing Research, c 2019 IEEE.

[30] A. Kazi, F. M. Mirkar, G. S. Patil, R. R. Kasar, *"Detecting E Banking Phishing Websites Using Associative Classification"*, International Journal of Engineering Technology Science and Research, ISSN 2394 – 3386, Volume 4, Issue 10, October 2017

[31] M. Kaytan, D. Hanbay, *"Effective Classification of Phishing Web Pages Based on New Rules by Using Extreme Learning Machines"*, Anatolian Journal of Computer Sciences Volume: 2 No: 1 2017 © Anatolian Science pp:15-36

[32] H. S. Hota, A. K. Shrivas, R. Hota, *"An Ensemble Model for Detecting Phishing Attack with Proposed Remove-Replace Feature Selection Technique"*, Procedia Computer Science 132 (2018) 900–907

[33] G. Vrbančič, I. FisterJr, V. Podgorelec, *"Swarm Intelligence Approaches for Parameter Setting of Deep Learning Neural Network: Case Study on Phishing Websites Classification"*, WIMS '18, June 25–27, 2018, Novi Sad, Serbia © 2018 ACM.

[34] N. Moradpoor, B. Clavie, B. Buchanan, *"Employing Machine Learning Techniques for Detection and Classification of Phishing Emails"*, Computing Conference 2017 18-20 July 2017 | London, UK.