WebSecure:

Phishing URLs Detection Using ML Techniques to Help Prevent Cyber Threats

by

Battalapalli Sai Divya (20BCE1970)

Emmay Koushal (20BAI1210)

Lakshmi Sumana N (20BAI1089)

A project report submitted to

Dr. Anusha K

Associate Professor, School of Computer Science and Engineering

in partial fulfilment of the requirements for the course of

CSE3501 – INFORMATION SECURITY ANALYSIS AND AUDIT

in

B. TECH., COMPUTER SCIENCE AND ENGINEERING

Vellore Institute of Technology, Chennai
Vandalur – Kelambakkam Road
Chennai – 600 127

November 2022

Chennai

BONAFIDE CERTIFICATE

This is to certify that the Project work titled **"WebSecure: Phishing URLs Detection Using ML Techniques to Help Prevent Cyber Threats"** that is being submitted by 20BCE1970, Battalapalli Sai Divya, 20BAI1210, Emmay Koushal, 20BAI1089 Lakshmi Sumana N is in partial fulfilment of the requirements for the award of **Bachelor of Technology in Computer Science and Engineering**, is a record of bonafide work done under my guidance. The contents of this Project work, in full or in parts, have neither been taken from any other source nor have been submitted to any other Institute or University for award of any degree or diploma and the same is certified.

Dr. ANUSHA K

Guide

ABSTRACT

Phishing is a form of social engineering assault that is frequently employed to obtain user information, such as login credentials and credit card details. It happens when an attacker deceives a victim into opening an email, instant message, or text message by disguising themselves as a reliable source. Here are several examples: As many faculty members as feasible receive a mass distribution of a counterfeit email purporting to be from myuniversity.edu. The user's password, according to the email, is soon to expire. The instructions state that they must renew their password within 24 hours by visiting myuniversity.edu/renewal. As a result, we provide in this work a thorough learning-based approach to deal with empowering high accuracy recognition of phishing destinations. To distinguish legitimate websites from phishing sites, the suggested method employs a pipeline model (a combination of Logistic Regression Classifier and Multinomial Naive Bayes Classifier) for high precision. Using a dataset of 5,49,346 entries, 72% of which are actual and 28% of which are phishing sites, we evaluate the models. According to the findings of thorough analyses, our pipeline-based technique proves to be incredibly effective in locating obscure phishing websites. Additionally, the pipeline-based methodology outperformed conventional individual machine learning classifiers evaluated on the same dataset, achieving a 98.03% phishing location rate with an F1-score of 0.97. The method described in this study compares favourably to the best in its field for phishing site identification based on deep learning.

Keywords:

Phishing, Deep Learning, Machine Learning, Phishing, Pipeline Model, Cyber Security, Social Engineering, Malware

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Phishing is sometimes referred to as a form of modernised mass fraud that preys on people's natural tendencies using the Internet to lure them in and steal their money.

It makes sense that in the past few years, phishing attacks have grown to pose a serious threat to global security. The main goal of these is to take advantage of the framework's shortcomings, which may be specialised or caused by clients' ignorance. Phishing is a real cybercrime that happens frequently.

Measurements show that 1 in every 99 communications is a phishing attack. According to Statics, the number of phishing attacks increased by 65% between 2016 and 2017.

In 2018, 83% of people experienced phishing attacks, which resulted in a variety of disruptions and harms. This includes a 67% decrease in income, a 54% loss of validity information, and a 50% decrease in property damage. The highlighted text or additional details make it clear that phishing is a challenging problem in numerous areas. The tool that can reduce this assault generally is AI, which can develop phishing techniques. We will prepare our model for the informational collection that contains the key features of the phishing sites using AI.

CHAPTER II

RELATED WORKS

2.1 Phishing Website Detection using Machine Learning: A Review

Author(s) - Purvi Pujara, M.B.Chaudhari

Year of Publication - 2019

Phishing is an approach to get a client's private data through email or site. As utilization of the web is extremely immense, practically everything is accessible online now it is either about looking for garments, electronic contraptions, earthenware or to pay for versatile, TV and power bills. Instead of hanging out in line for quite a long time, individuals are monitoring and utilizing web strategies. Because of this phisher has a wide extension to execute phishing tricks. As there is a ton of exploration work done here, there isn't any single method, which is sufficient to identify a wide range of phishing assault. As innovation increases, phishing assaults utilize new strategies step by step. This empowers us to discover successful classifiers to identify phishing. In this paper, the creators played out an itemized writing study about phishing site discovery. According to this, they reasoned that tree-based classifiers in AI approach is preferable over others.

2.2 **Detection of Phishing Attacks with Machine Learning Techniques in Cognitive Security Architecture**

Author - Roberto O. Andrade, and Mar´ıa Cazares

Year of Publication - 2019

A typical conduct is a typical issue as found in this paper. In this record the creators experience these regular issues and dissected them, it is because of these investigates and results that the creators can put down these focuses:

• Artificial Intelligence is a decent instrument to confront this abnormal conduct, since it is quicker, is effective and current innovation lets us grow better applications.

• Some phishing strategies like shortening URLs could be confronted with devices like this AI application that is skilled to decide whether a URL are positive or negative, the following stage after that is add this URL to boycott.

• Although this AI couldn't be consistently correct, it very well may have registered those URLs with a web checker of shortening URLs. Which carries us to the following proposal.

• Is recommendable not to open an abbreviated URL prior to checking it, since it is as of now realized that a significant number of those URLs could be phishing assaults, malware, etc.

• By knowing these shortcomings, as a network it is conceivable to attempt to grow new devices that help us to improve these frail procedures.

2.3 **Deep Learning for Phishing Detection**

Author - Yao,Yuan Hao Ding, Xiaoyong Li

Year of Publication - 2018

Because of the possible danger of phishing assaults in two dimensional code, this paper proposes an improved FPN joined with Faster R-CNN logo acknowledgment technique. In light of this, character consistency is utilized to pass judgment on the legitimacy of the two-dimensional code. The viability of our proposed strategy is demonstrated by contrasting other logo acknowledgment strategies and phishing location techniques. Furthermore, they accept that since the little measured logo picture influences the acknowledgment strategy, if the super-goal amplification of the little article can be proceeded quite far without bending, the impact on the acknowledgment technique will be decreased. The creators expressed that they will investigate picture super goals in future work.

2.4 PhishLimiter: A Phishing Detection and Mitigation Approach Using Software-Defined Networking

Author - Chengbin Hu

Year of Publication - 2017

A phishing assault is an extremely basic social designing way to deal with focusing on an association and end clients. It has gotten one of the most destructive assaults these days. There have been various investigations on identifying and moderating phishing assaults. Conventional arrangements centre around the utilization of inline review procedures, for example, an IPS or intermediary administration dependent on static string coordinating in customary IDS, for example, SNORT and BRO. In this paper, they have proposed PhishLimiter as another answer for defeat phishing assaults. Phish Limiter can deal with network traffic elements for containing phishing assaults and can give a superior traffic the executives since it has a worldwide perspective on networks because of SDN. They have first arranged phishing marks by building up an ANN model utilizing a PLS framework. They at that point have assembled a dependable framework for SDN traffic stream designing by presenting PLS and the OVS exchanging score for SF and FI modes that can utilize the programmability of SDN to manage the elements of phishing assaults in reality. The reliable framework has been accomplished through our constructed crypto module for DPI through SSL/RSA encoded traffic. Besides, they have hypothetically and tentatively assessed of Phish Limiter. They have assessed the reliability of each SDN stream to distinguish any potential

dangers dependent on every profound bundle investigation. Similarly, they have seen how the proposed assessment approach of two SF and FI modes inside PhishLimiter recognizes and mitigates phishing assaults prior to arriving at end clients if the stream has been resolved conniving. Utilizing our real-world test assessment on GENI and phishing dataset, they have shown that Phish Limiter is a compelling and proficient answer for identifying and moderate phishing assaults with its exactness of 98.39%.

2.5 A Phishing Detection System based on Machine Learning

Author - Che-Yu Wu, Cheng-Chung Kuo, Chu-Sing Yang

Year of Publication - 2019

As the Internet has become a basic piece of individuals' carries on with, a developing number of individuals are appreciating the comfort brought by the Internet, while more are assaults originating from on the clouded side of the Internet. In light of certain shortcomings of human instinct, programmers have planned confounding phishing pages to tempt web watchers to proactively uncover their security, touchy data. In this article, the writers propose a URL based location framework - consolidating the URL of the website page URL and the URL of the page source code as highlights, import Levenshtein Distance as the calculation for computing the similitude of strings, and enhanced by the AI engineering. Because of the extraordinary precision in little example numbers and double arrangement, they actualize Support-vector machines to be the AI calculation model in our framework. The framework is intended to give high precision and low bogus positive rate discovery results for obscure phishing pages. As the Internet has become a basic portion of individuals' carries on with, a developing number of individuals are getting a charge out of the accommodation brought by the Internet, while more are assaults originating from on the clouded side of the Internet. Considering certain shortcomings of human instinct, programmers have planned confounding phishing pages to allure web watchers to proactively uncover their protection, touchy data. In this article, the writers propose a URL-based identification framework - joining the URL of the site page URL and the URL of the website page source code as highlights, import Levenshtein Distance as the calculation for figuring the likeness of strings, and enhanced by the AI design. Because of the incredible precision in little example numbers and twofold arrangement, they actualize Support-vector machines to be the AI calculation model in our framework. The framework is intended to give high exactness and low bogus positive rate location results for obscure phishing pages.

2.6 PhishMon: A Machine Learning Framework for Detecting Phishing Webpages

Author - Amir Reza Niakan Lahiji, Bei-Tseng Chu, Ehab Al-Shaer

Year Of Publication - 2016

In this paper, the creators proposed PhishMon, a component rich AI framework to distinguish phishing sites. It depends on eighteen notable highlights, fifteen of which are new, to choose whether the page pointed by a given URL is a phish. These highlights can be productively determined, without requiring connection with any outsider framework, which can

restrictively defer the dynamic cycle. They catch different attributes of the web application and its fundamental framework. By utilizing these highlights, they by implication measure the measure of exertion put into advancement and arrangement of a web application, which is astoundingly extraordinary among genuine and phishing sites. They additionally directed broad examinations on a genuine world dataset containing affirmed phishing and real examples gathered from the Internet between Sept. also, Nov. 2017. Our outcomes show that PhishMon accomplishes a serious extent of precision in recognizing authentic pages from concealed phishing pages without raising numerous bogus alerts: on the gathered dataset, PhishMon arrives at a 95.4% identification rate with 1.3% bogus positive.

2.7 A Novel Machine Learning Approach to Detect Phishing Websites

Author - Ishant Tyagi, Jatin Shad, Shubham Sharma, Siddharth Gaur, Gagandeep Kaur,

Year of Publication - 2018

This paper for the most part contains AI procedures to distinguish the phishing sites. Phishing sites generally recover client's data through login pages. They are predominantly inspired by the bank subtleties of the clients. Out of the numerous highlights considered, the main one was HTTPS with SSL for example regardless of whether a site utilizes HTTPS, backer of authentication is trusted or not, and the time of testament ought to be in any event one year. Later on, they might want to broaden our undertaking by making an augmentation to impede the recognized phishing site at whatever point the client taps on their connection.

2.8 Detecting Phishing Websites Using Machine Learning

Author - Amani Alswailem, Norah Alrumayh, Bashayr Alabdullah, Dr.Aram Al Sedrani

Year of Publication - 2018

Phishing site is one of the web securities issues that target human weaknesses as opposed to programming weaknesses. It very well may be portrayed as the way toward drawing in online clients to acquire their touchy data, for example, usernames and passwords. In this paper, they offer an astute framework for identifying phishing sites. The framework goes about as an extra usefulness to a web program as an expansion that consequently tells the client when it distinguishes a phishing site. The framework depends on an AI strategy, especially regulated learning. They have chosen the Random Forest strategy because of its great exhibition in characterization. Their center is to seek after a better classifier by examining the highlights of a phishing site and pick the better blend of them to prepare the classifier. Accordingly, they closed our paper with a precision of 98.8% and a mix of 26 highlights.

2.9 Malicious URL Detection using Machine Learning: A Survey

Author - DOYEN SAHOO, CHENGHAO LIU, STEVEN C.H. HOI

Year of Publication - 2017

Vindictive URL, a.k.a. a malignant site, is a typical and genuine danger to network protection. Malignant URLs have spontaneous substance (spam, phishing, drive-by downloads, and so forth) and bait clueless clients to become survivors of tricks (financial misfortune, burglary of private data, and malware establishment), and cause misfortunes of billions of dollars consistently. It is basic to distinguish and follow up on such dangers in an ideal way. Customarily, this recognition is done generally through the utilization of boycotts. Notwithstanding, boycotts can't be comprehensive, and come up short on the capacity to identify recently created malignant URLs. To improve the consensus of malignant URL locators, AI procedures have been investigated with expanding consideration as of late. This article expects to give an extensive review and an auxiliary comprehension of Malicious URL Detection procedures utilizing AI. They present the proper definition of Malicious URL Detection as an AI task, and order and audit the commitments of writing considers that tends to various elements of this issue (highlight portrayal, calculation plan, and so on) Further, this article gives an ideal and far reaching review for a scope of various crowds, not just for AI scientists and specialists in the scholarly world, yet in addition for experts and professionals in the online protection industry, to assist them with understanding the cutting edge and encourage their own exploration and reasonable applications. They likewise talk about handy issues in framework configuration, open exploration difficulties, and point out significant bearings for future examination.

2.10 Intelligent phishing website detection using random forest classifier

Author - Subasi, E. Molah, F. Almkallawi and T. J. Chaudhery

Year of Publication - 2017

Phishing is characterized as emulating a respectable organization's site intending to take private data of a client. To kill phishing, various arrangements were proposed. Be that as it may, just one single sorcery projectile cannot wipe out this danger totally. Information mining is a promising procedure used to identify phishing assaults. In this paper, a smart framework to distinguish phishing assaults is introduced. They utilized diverse information mining procedures to choose classes of sites: genuine or phishing. Various classifiers were utilized so as to develop an exact keen framework for phishing site location. Grouping exactness, zone under beneficiary working trademark (ROC) bends (AUC) and F-measure is utilized to assess the exhibition of the information mining methods. Results demonstrated that Random Forest has outflanked best among the order strategies by accomplishing the most noteworthy precision 97.36%. Irregular woods runtimes are very quick, and it can manage various sites for phishing identification.

2.11 Detection of Phishing Websites using an Efficient Machine Learning Framework

Author - Naresh Kumar D , Nemala Sai Rama Hemanth , Premnath S , Nishanth Kumar V, Uma S

Year of Publication - 2020

Phishing assault is one of the normally known assaults where the data from the web clients are taken by the gate crasher. The web clients lose their touchy data, for example, Protected passwords, individual data and their exchanges to the gate crashers. The Phishing assault is regularly conveyed by the aggressors where the genuine as often as possible utilized sites are controlled and concealed to accumulate the individual data of the clients. The Intruders utilize the individual data and can control the exchanges and get unmistakable from them. From the writing there are different enemies of Phishing sites by the different writers. A portion of the strategies are Blacklist or Whitelist and heuristic and visual closeness-based techniques. Regardless of the clients utilizing these procedures most of the clients are getting assaulted by the gate crashers by methods for Phishing to assemble their delicate data. A tale Machine Learning based arrangement calculation has been proposed in this paper which utilizes heuristic highlights where include choice can be separated from the properties, for example, Uniform Resource Locator, Source Code, Session, Type of security include, Protocol utilized, kind of site. The proposed model has been assessed utilizing five AI calculations, for example, irregular backwoods, K Nearest Neighbour, Decision Tree, Support Vector Machine, Logistic relapse. Out of these models, the arbitrary timberland calculation performs better with assault identification precision of 91.4%. The Random Forest Model uses symmetrical and angled classifiers to choose the best classifiers for precise identification of Phishing assaults in the sites.

CHAPTER III

3. Proposed Method/ System Design

3.1 Structure/Modules:

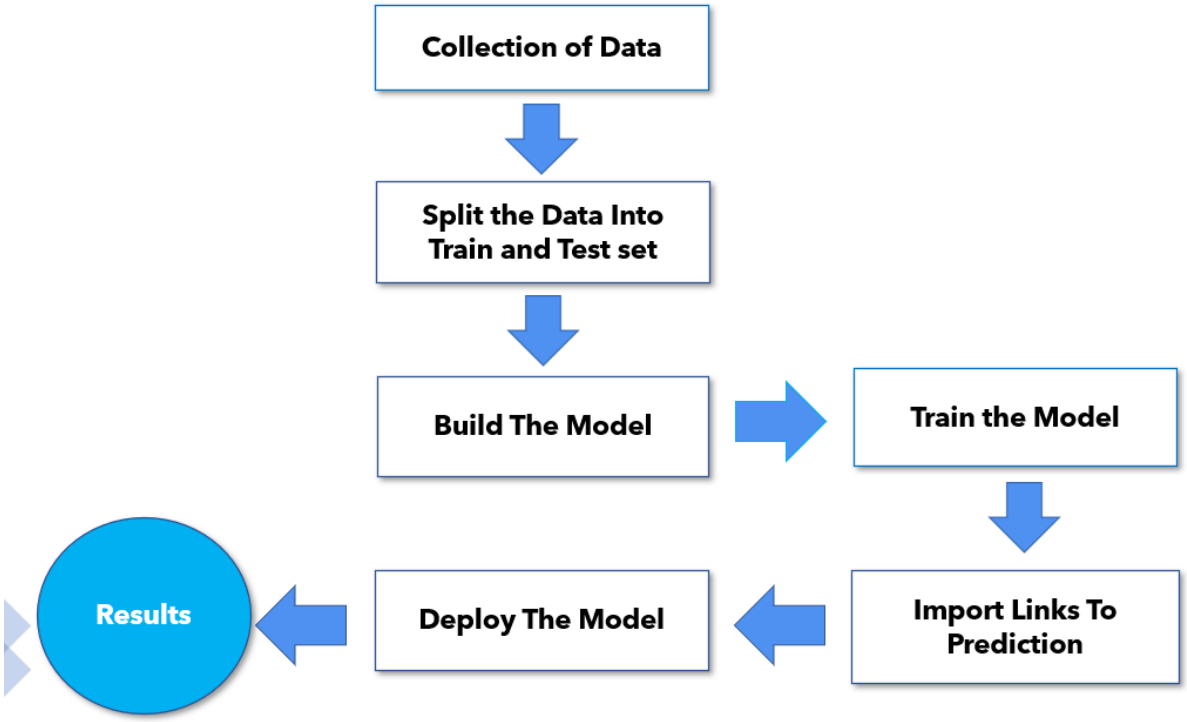

Figure 1: Worklfow

- Machine learning module (sklearn)
- logistics regression
- pickel module
- imaplib library
- Email library
- Interface: Graphical interface tkinter python module

3.2 Dataset

https://www.kaggle.com/datasets/taruntiwarihp/phishing-site-urls

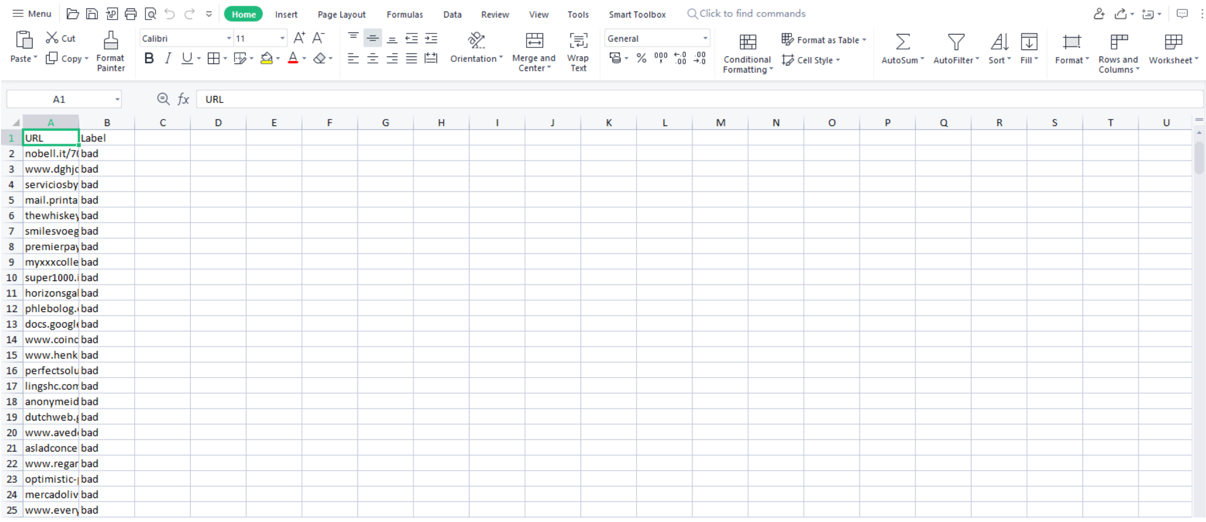

Table 1: Data csv file

- Data is containg 5,49,346 entries.
- There are two columns.
- Label column is prediction col which has 2 categories
  A. Good - which means the URLs is not containing malicious stuff and this site is not a Phishing Site.
  B. Bad - which means the URLs contains malicious stuff and this site is a Phishing Site.
- There is no missing value in the dataset.

```
phish_data.head()
```

	URL	Label
0	nobell.it/70ffb52d079109dca5664cce6f317373782/...	bad
1	www.dghjdgf.com/paypal.co.uk/cycgi-bin/webscrc...	bad
2	serviciosbys.com/paypal.cgi.bin.get-into.herf....	bad
3	mail.printakid.com/www.online.americanexpress....	bad
4	thewhiskeydregs.com/wp-content/themes/widescre...	bad

Table 2: Dataset

3.3 Data Visualization:

```
#visualizing target_col
sns.set_style('darkgrid')
sns.barplot(label_counts.index,label_counts.Label)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fe5105474d0>
```

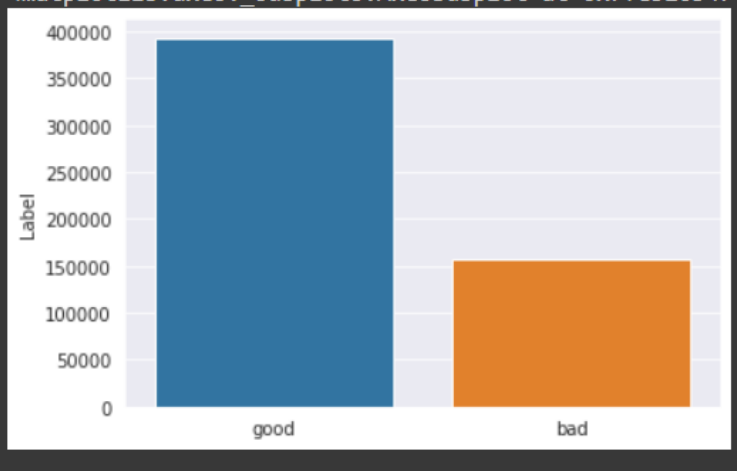

Figure 2: Dataset as a bar graph

3.4 Regex Tokenizer:

Tokenization, a pre-processing procedure that produces a list of tokens from text, converts the tokens into one-hot vectors that can be fed into transformers or bi-directional LSTM models for NLP tasks including machine translation, summarization, sentiment analysis, and coreference resolution, to name a few. (After the training procedure, each token can be given an n-dimensional vector embedding that captures a particular meaning.)

The URLs were then tokenized to produce vectors of useful keywords.

```
phish_data.sample(5)
```

	URL	Label	text_tokenized
515170	91.239.24.22:6892	bad	[]
257586	web1.ctaa.org/webmodules/webarticles/anmviewer...	good	[web, ctaa, org, webmodules, webarticles, anmv...
263393	123people.ca/s/sylvie+ferland	good	[people, ca, s, sylvie, ferland]
214456	monacoradio.com/	good	[monacoradio, com]
269854	aeispeakers.com/speakerbio.php?SpeakerID=588	good	[aeispeakers, com, speakerbio, php, SpeakerID]

Table 3: After Tokenization

3.5 Snowball Stemmer

This stemming algorithm, commonly referred to as the Porter2 stemming algorithm, is an improved version of the Porter Stemmer because some of its shortcomings have been addressed.

Stemming is the process of stripping a word down to its root, or lemma, which attaches to suffixes, prefixes, and other word parts. In plain English, stemming is the process of reducing a word to its root word or stem so that terms of the same kind are grouped together under a single stem. As an illustration, the words care, cared, and caring all share the same root word. In the processing of natural language, stemming is crucial.

```
phish_data.sample(5)
```

	URL	Label	text_tokenized	text_stemmed
271703	allcdcovers.com/show/78576/celine_dion_a_new_d...	good	[allcdcovers, com, show, celine, dion, a, new,...	[allcdcov, com, show, celin, dion, a, new, day...
65216	www.bestwebimage.com/web-services-and-tools/tw...	good	[www, bestwebimage, com, web, services, and, t...	[www, bestwebimag, com, web, servic, and, tool...
328015	facebook.com/mchllkinney	good	[facebook, com, mchllkinney]	[facebook, com, mchllkinney]
521872	bluebit.ga/UF4IBxpJUl8VHldEBUsGDhNdVF4QTAFZSgF...	bad	[bluebit, ga, UF, IBxpJUl, VHldEBUsGDhNdVF, QT...	[bluebit, ga, uf, ibxpjul, vhldebusgdhndvf, qt...
421197	redstormsports.com/sports/m-baskbl/spec-rel/03...	good	[redstormsports, com, sports, m, baskbl, spec,...	[redstormsport, com, sport, m, baskbl, spec, r...

Table 4: After Stemming

3.6 Count Vectorizer:

Text data can be used directly in machine learning and deep learning models, including text categorization, thanks to Countvectorizer. Characters and words are not understood by machines. So, for a machine to understand text data, it must be represented in numerical form. Text can be transformed into numerical data with the Countvectorizer tool.

```
phish_data.sample(5)
```

	URL	Label	text_tokenized	text_stemmed	text_sent
445645	themarknews.com/articles/3159-the-cost-of-clos...	good	[themarknews, com, articles, the, cost, of, cl...	[themarknew, com, articl, the, cost, of, close...	themarknew com articl the cost of close canadi...
52182	sites.google.com/a/bestechsales.com/bestech-sa...	good	[sites, google, com, a, bestechsales, com, bes...	[site, googl, com, a, bestechsal, com, bestech...	site googl com a bestechsal com bestech sale
476272	youtube.com/watch?v=fzzGYPDbEaY	good	[youtube, com, watch, v, fzzGYPDbEaY]	[youtub, com, watch, v, fzzgypdbeay]	youtub com watch v fzzgypdbeay
506061	cignitech.com/076wc?mhbvhpjyqyp=cuseks	bad	[cignitech, com, wc, mhbvhpjyqyp, cuseks]	[cignitech, com, wc, mhbvhpjyqyp, cusek]	cignitech com wc mhbvhpjyqyp cusek
370069	lansing.lib.il.us/catalog.html	good	[lansing, lib, il, us, catalog, html]	[lans, lib, il, us, catalog, html]	lans lib il us catalog html

Table 5: After Count Vectorization

CHAPTER IV

SIMULATION / IMPLEMENTATION RESULTS

4.1 Logistic Regression Report:

```
Training Accuracy : 0.979165503666182
Testing Accuracy : 0.964605313935793

CLASSIFICATION REPORT

              precision    recall  f1-score   support

         Bad       0.91      0.97      0.94     36694
        Good       0.99      0.96      0.98    100643

    accuracy                           0.96    137337
   macro avg       0.95      0.97      0.96    137337
weighted avg       0.97      0.96      0.96    137337


CONFUSION MATRIX
<matplotlib.axes._subplots.AxesSubplot at 0x7fe4d7cb5ed0>
```

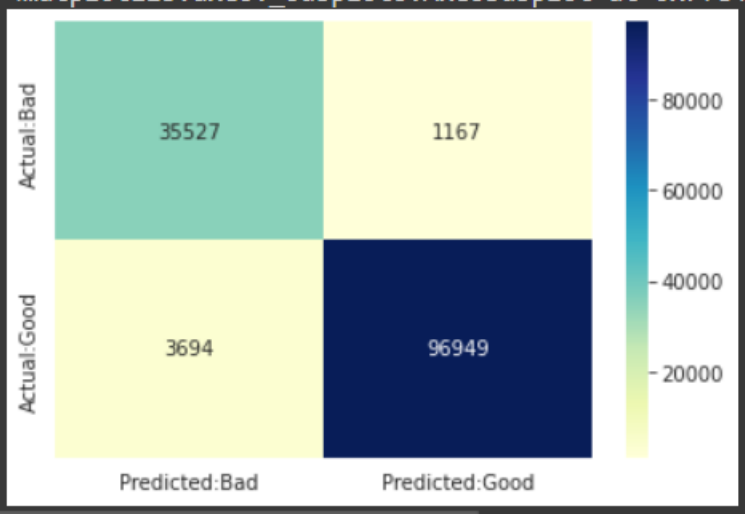

Figure 3: LR Report

4.2 Multinomial Naïve Bayes Report:

```
Training Accuracy : 0.97408794468082
Testing Accuracy : 0.957622490661657

CLASSIFICATION REPORT

              precision    recall  f1-score   support

         Bad       0.91      0.94      0.92     38283
        Good       0.98      0.97      0.97     99054

    accuracy                           0.96    137337
   macro avg       0.94      0.95      0.95    137337
weighted avg       0.96      0.96      0.96    137337


CONFUSION MATRIX
<matplotlib.axes._subplots.AxesSubplot at 0x7fe4d7cc7e10>
```

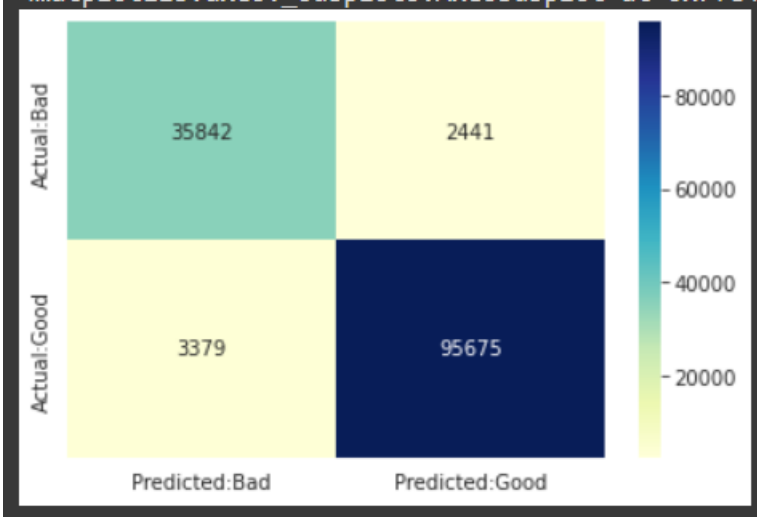

Figure 4: Mnb Report

4.3 Comparison:

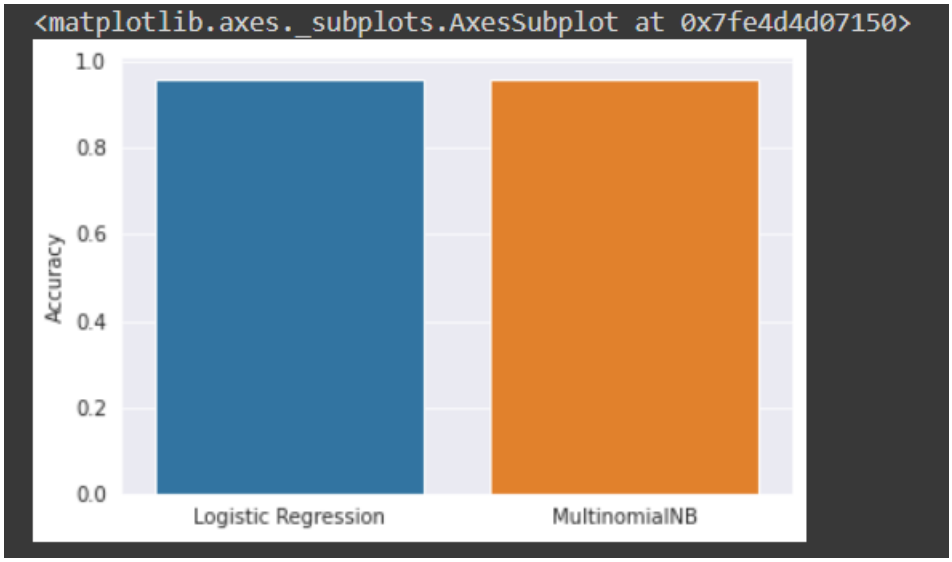

Figure 5: Comparison Plot

4.4 Pipeline Classifier Report:

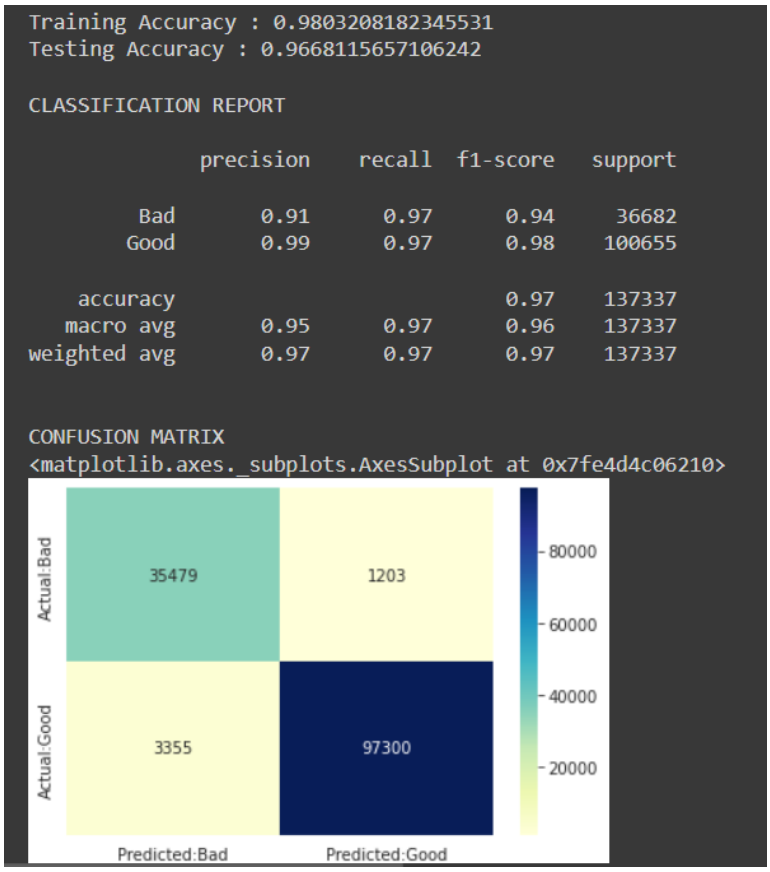

Figure 6: Pipeline Model Report

4.5 Home Page of the Website:

Figure 7: Homepage

4.6 Entering a malicious URL:

Figure 8: Malicious URL Demo

Output:

Figure 9: Malicious URL Output

4.7 Entering a safe URL:

Figure 10: Safe URL Demo

Output:

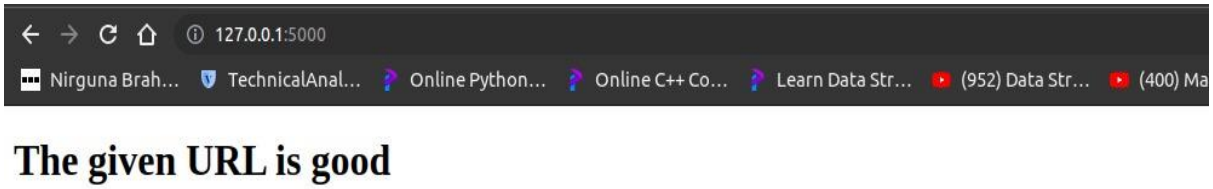

Figure 11: Safe URL Output

4.8 Enabling the IMAP access in Gmail:

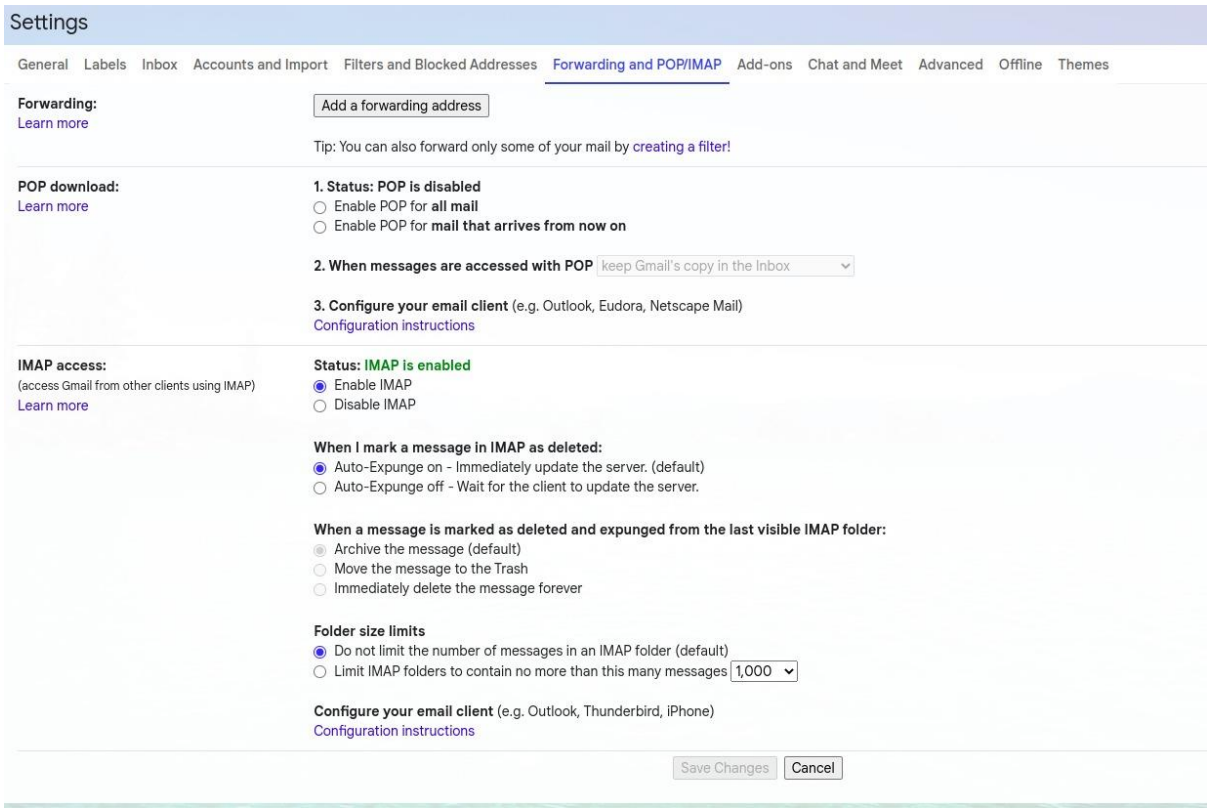

Figure 12: IMAP Access

4.9 Select device and generate password:

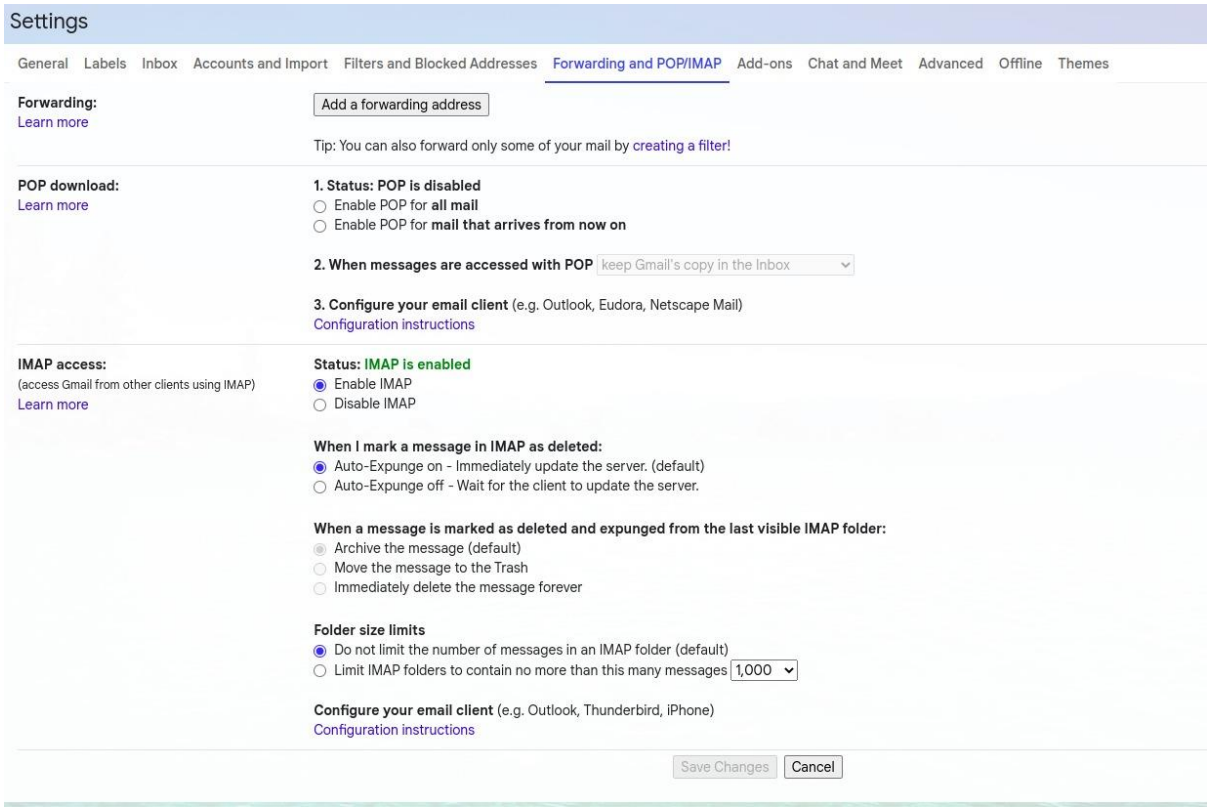

Figure 13: Password Generation

4.10 Working of Tool:

Malicious URL Detection from Emails:

Figure 14: Detection Tool

Output:

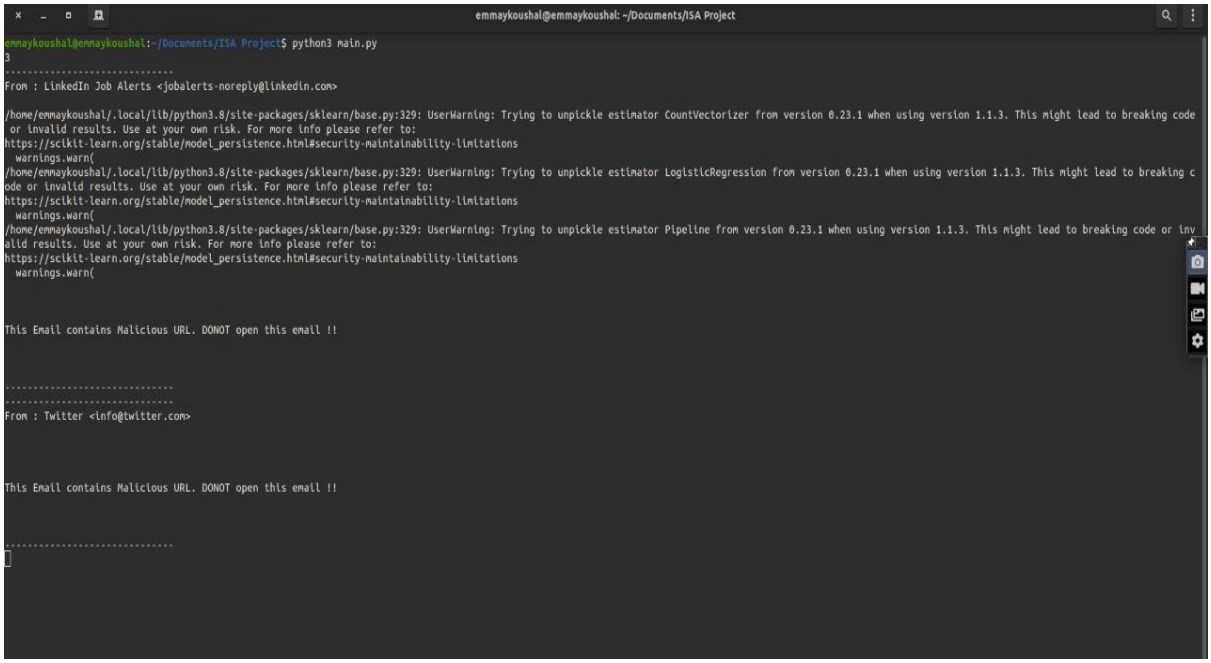

Figure 15: Output

CHAPTER V

CONCLUSION AND FUTURE WORKS

5.1 Conclusion

In this project, we proposed a machine learning method based on pipeline models for identifying phishing websites. On a benchmarked dataset of 5,49,346 distinct URLs from phishing sites and legitimate websites, we thoroughly examined the model. The model outperforms a few common AI classifiers tested on the same dataset. The results show that our suggested pipeline-based approach can be used to identify new, already obvious phishing sites than other methods more accurately. In our upcoming work, we intend to strengthen the model preparation process by computerising the research and identification of the key influencing boundaries that together result in the ideal performance of the model.

5.2 Future works

1. Web Application/ Extension:
Let us face it, in today's world, we all expect a mobile app for any cool service. As a result, another future enhancement could be the creation of an App with the same features as Web Applications.

2. Work on improving the accuracy and extend the model agility by applying it to various other datasets.

3. Launch the trained model as an open-source pre-trained model that can be used for transfer learning for other problem statements.

References

[1] Alswailem, A., Alabdullah, B., Alrumayh, N., & Alsedrani, A. (2019, May). Detecting Phishing Websites Using Machine Learning. In 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS) (pp. 1-6). IEEE.

[2] Alkawaz, M. H., Steven, S. J., & Hajamydeen, A. I. (2020, February). Detecting Phishing Website Using Machine Learning. In 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA) (pp. 111-114). IEEE.

[3] Patil, V., Thakkar, P., Shah, C., Bhat, T., & Godse, S. P. (2018, August). Detection and Prevention of Phishing Websites Using Machine Learning Approach. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-5). IEEE.

[4] James, J., Sandhya, L., & Thomas, C. (2013, December). Detection of phishing URLs using machine learning techniques. In 2013 international conference on control communication and computing (ICCC) (pp. 304-309). IEEE.

[5] Abdelhamid, N., Thabtah, F., & Abdel-jaber, H. (2017, July). Phishing detection: A recent intelligent machine learning comparison based on models content and features. In 2017 IEEE international conference on intelligence and security informatics (ISI) (pp. 72-77). IEEE.

[6] Kumar, J., Santhanavijayan, A., Janet, B., Rajendran, B., & Bindhumadhava, B. S. (2020, January). Phishing Website Classification and Detection Using Machine Learning. In 2020

International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-6). IEEE.

[7] Tyagi, I., Shad, J., Sharma, S., Gaur, S., & Kaur, G. (2018, February). A novel machine learning approach to detect phishing websites. In 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 425-430). IEEE.

[8] Niakanlahiji, A., Chu, B. T., & Al-Shaer, E. (2018, November). PhishMon: a machine learning framework for detecting phishing webpages. In 2018 IEEE International Conference on Intelligence and Security Informatics (ISI) (pp. 220-225). IEEE.

[9] Vilas, M. M., Ghansham, K. P., Jaypralash, S. P., & Shila, P. (2019, December). Detection of Phishing Website Using Machine Learning Approach. In 2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT) (pp. 384-389). IEEE.

[10] Wu, C. Y., Kuo, C. C., & Yang, C. S. (2019, August). A Phishing Detection System based on Machine Learning. In 2019 International Conference on Intelligent Computing and its Emerging Applications (ICEA) (pp. 28-32). IEEE.

[11] Jain, A. K., & Gupta, B. B. (2016, March). Comparative analysis of features based machine learning approaches for phishing detection. In 2016 3rd international conference on computing for sustainable global development (INDIACom) (pp. 2125-2130). IEEE.

[12] Chin, T., Xiong, K., & Hu, C. (2018). Phishlimiter: A phishing detection and mitigation approach using software-defined networking. IEEE Access, 6, 42516-42531.

[13] Sadique, F., Kaul, R., Badsha, S., & Sengupta, S. (2020, January). An Automated Framework for Real-time Phishing URL

Detection. In 2020 10th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0335-0341). IEEE.

[14] Yao, W., Ding, Y., & Li, X. (2018, December). Deep learning for phishing detection. In 2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom) (pp. 645-650). IEEE.

[15] Garcés, I. O., Cazares, M. F., & Andrade, R. O. (2019, December). Detection of Phishing Attacks with Machine Learning Techniques in Cognitive Security Architecture. In 2019 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 366-370). IEEE