EXPLORATORY DATA ANALYSIS OF PHISHING SITES TO IDENTIFY MOST IMPORTANT FEATURES TO DETECT A PHISHING SITE

BY

MOST. SABINA YASMIN

162-15-7680

AND

MOU ROY

162-15-7677

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

MD. SADEKUR RAHMAN

Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Ms. FARAH SHARMIN

Senior Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

OCTOBER 2020

APPROVAL

This Project titled "**Exploratory data analysis of phishing sites to identify most important features to detect a phishing site**", submitted by Most. Sabina Yasmin, ID: 162-15-7680 and Mou Roy, ID: 162-15-7677 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 7thOctober, 2020.

BOARD OF EXAMINERS

____

Dr. Sayed Akhter Hossain **Chairman**
Professor and Head
Department of Computer and Science Engineering
Faculty of Science & Information Technology
Daffodil International University

____

Md. Zahid Hasan **Internal Examiner**
Assistant Professor
Department of Computer and Science Engineering
Faculty of Science & Information Technology
Daffodil International University

____

Most. Hasna Hena **Internal Examiner**
Assistant Professor
Department of Computer and Science Engineering
Faculty of Science & Information Technology
Daffodil International University

____

Dr. Mohammad Shorif Uddin **External Examiner**
Professor
Department of Computer and Science Engineering
Jahangirnagar University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Mr. Sadekur Rahman, Assistant Professor** and **Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Md. Sadekur Rahman
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Yasmin

Most. Sabina Yasmin
162-15-7680
Department of CSE
Daffodil International University

Mou Roy
162-15-7677
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Md. Sadekur Rahman**, **Assistant Professor**, Department of CSE, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of Computer Science inspired us to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Syed Akhter Hossain**, **Professor**, **and Head,** Department of Computer Science and Engineering, Daffodil International University for his kind help to finish our project and also to other faculty members and the staff of Computer Science and Engineering department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Phishing is one of the top most cybercrime according to a lot cybercrime awareness organization. “**Exploratory data analysis of phishing sites to identify most important features to detect a phishing site**" is a research project which aims to explore the most significant features of a phishing site in order to detect a phishing site. In order to explore these features data were collected from an open source machine learning data repository. Later correlation and univariate selection methods were applied to discover the most significant features to detect a phishing site. Finally, based on the top five selected features a system was built to check whether it can identity phishing sites or not.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

In the present globe, it is beyond thoughts to go without technologies such as the internet, mobile phone, computers and others. It is quite unimaginable for them to go a day without the presence of technology. But sometimes we also face cybercrime through technology. Phishing is a type of social engineering attack. Phishing is a method of trying to gather personal information using deceptive emails and websites. It is a fraudulent attempt to obtain personal information or data such as usernames, passwords and credit card details, by disguising oneself as a trustworthy entity in an electronic communication. Typically, it is carried out by email spoofing, instant messaging and text messaging. Users directly enter their personal information at a fake website which matches the look and feel of the legitimate site.

Without the help of social networking, we are not able to communicate with our friends, relatives and others easily. We can’t completely eliminate phishing attacks. But we can protect people from the harm of phishing attacks by taking some steps. Further we will discuss the way to identify phishing websites by testing algorithms and check the accuracy of these websites. Then we can build some efficient steps to secure our personal information.

1.2 Motivation

The motivation behind the phishing attack is no different than any other information security incident. Generally, attackers will be looking to trick the target user into divulging credentials on a pharming website.

The primary motivation for hackers is the money they can obtain by stealing our password, bank details, holding the customer information or selling your data to competitors or on the dark web.

Attackers try to steal consumer’s personal information. When a user opens a fake web page and enters the username and protected password, the credentials of the user are acquired by the attacker which can be used for malicious purposes. Phishing websites look very similar in appearance to their corresponding legitimate websites to attack large number of Internet user. Recent developments in phishing detection have led to the growth of numerous new visual similarity-based approaches. The fake website is the clone of targeted genuine website and it always contains some input fields. When the user submits his/her personal details, the information is transferred to the attacker. An attacker steals the credential of the innocent user by performing following steps.

Construction of phishing site:

In the first step the attacker identifies the target as a well-known organization. By visiting their website. The attacker then uses their information to construct the fake website.

URL sending:

In this step, the attacker composes a bogus email and sends it to the thousands of users. Attacker attacked the URL of the user of the fake website in the bogus e-mail.

Attacker uses this credential for malicious purposes. For example, attackers purchase something by using credit card details of the user. We proposed that there are some ways in which the solution to phishing can be approached. Detect phishing attacks, before they reach the user. Default once the user has reached the phishing site. But the best method is an approach utilizing a mix of all threes. Phishing is evolving day to day to avoid detection and by taking on all and we increase the chances that they will be found and stopped.

1.3 Rationale of the Research

Phishing is a common threat for all kinds of internet users. It is ranked among the 5 top most common cyber threats by different organizations [8,9]. Therefore, studying the features of these sites are very essential in order to facilitate other system builders to develop systems those can detect and prevent phishing.

14. Research Questions

In order to conduct the research, the following research questions have been set:

- Does correlation and univariate selection help to identify most significant features of a dataset?
- What are the most significant features of a phishing website?

1.5 Report Layout

Chapter 1 describes about what we are going to do in this research.

Chapter 2 summarizes the related works regarding our research identifies the scope of our research.

Chapter 3 narrates fundamental ideas and methodology behind our research.

Chapter 4 describes the experimental results of our research.

Chapter 5 talks about the ethical issues and social impact of our research.

Chapter 6 draws conclusion to our report.

CHAPTER 2

BACKGROUND

2.1 Introduction

As phishing is a common cybercrime for a long time, quite a number of researches have been done in this domain. In this chapter we have tried to summarize a few of those in order to identify the scope of our research.

2.2 Related works

- Roopak. S and Tony Thomas suggested a method based on the HTML source code matching. Here similar web pages are searched by Google and compare their HTML source code [1].
- Maher Aburrous, M. A. Hossain, FadiThabatah and KeshavDahal used the fuzzy logic technique for detecting phishing sites. Their proposed model is based on FL operators which are used to characterize the website phishing factors and indicators as fuzzy variables and produces six measures and criterions of phishing attack dimensions with a layer structure[2].
- Guang-Gang Geng, Xiao-Dong Lee, Wei Wang and Shian-Shyong Tseng proposed a method that is based on the favicon of a website. According to them, most phishers use custom favicons to trick the users, and they target almost all traditional industries including financial institutions, online payment services, insurance companies, governments, multiplayer games, email services, hotels, security services, social network sites, retail services, and auctions, etc. They used favicon detection and recognition methods to filter phishing websites [3].
- Luong Anh Tuan Nguyen, Lam To, HuuKhuong Nguyen and Minh Hoang Nguyen proposed a method based on the single-layer neural network. This proposed technique calculates the value of heuristics objectively. Then, the weights of heuristic are generated by a single-layer neural network [4].

- Ying Pan and Xuhua Ding intended an anomaly based phishing detection method. According to their method when a phishing site maliciously claims a false identity, it always demonstrates abnormal behaviors compared to an honest site which is indicated by some web DOM objects in the page and HTTP transaction and by capturing these anomalies phish sites can be detected [5].

- Sadia Afroz and Rachel Greenstadt presented a method to detect phishing attacks based on profiles of sensitive sites' appearance and content. Based on their study, this approach was able to identify phishing webpages using URL, HTML based contents and displayed images[6].

- Shraddha Parekh, Dhwanil Parikh, Srushti Kotak and SmitaSankhe proposed a new method to identify phishing websites by URL analysis. First they define some features of phishing webpages URLs, then match the URL contents with the selected website and confirm the identification of phishing[7].

2.3 Research Summary

Summary of the findings of our study is presented in table 2.1.

Table 2.1: Research summary

Authors	Features	Algorithm	Accuracy
Sadia Afroz and Rachel Greenstadt	length of the URL, the number of dots, host feature of the URL, include of IP address, DNS properties such as TTL and geographical location,	Contrast Context Histogram(CCH), K-mean algorithm, Scale Invariant Feature Transform(SIFT) algorithm, Knuth-Morris-part string	95-99%

	HTML,Javascript	search algorithm, Current image matching, Image segmentation, OCR algorithm	
Shraddha Poarekh, Dhwaril Parikh, Srushti Kotak and Prof. SmitaSankhe	Address Bar based Features, Abnormal Based Features, HTML and javascript Based Feature, Domain based Features	Classification algorithm, Search Method algorithm	95%
Ying Pan and Xuhua Ding	URL string[4,28,39], HTML[7,21], ‘’HTTPS’’, The structural ,lexical features, Host-based features, Email,website,URL, and social media features	TF-IDF algorithm, J48 algorithm, svm algorithm, Naive Bayesian algorithm, Keyword extraction algorithm	82%
Roopak.S and Tony Thomas	Mining URL by Google	HTML code comparison	87%

	search,Compare the pages,HTML source code comparison method, Comparison based on cosine similarity.	Algorithm,TF-IDF information retrieval algorithm	
Maher Aburrous, M.A. Hossain, FadiThabatah and KeshavDahal	URL & Domain identity,Security&Encryption,Source code &Javascript,Page style and contents,Web address Bar,Social Human Factor	AprioriAlgorithm,Fuzzy Data Mining Algorithm, Decision Tree(c4.5) Algorithm,PARTAlgorithm,JRip Ripper Algorithm,PRISM Algorithm	72%
Luong Anh Tuan Nguyen, Lam To, HuuKhuong Nguyen and Minh Hoang Nguyen	Bad form,Bad action field,Non-matching URLSs,Page in top search results,Search copyright brand plus domain, Search copyright brand plus hostname	TF-IDF information retrieval algorithm, Network Training Algorithm	99%

2.4 Scope of the problem

In depth study of the related works reveal that a lot of works have been done to study the URL of phishing websites to identify them as phishing sites. However, no works have been done on to study the most significant features of phishing sites. There is a huge scope to study these features and no doubt it will help professionals to detect phishing websites.

2.5 Challenges

Lack of knowledge about phishing sites were the most challenging thing for us to carry out this research at the very beginning. Then collecting dataset and research experience become another issue those puled us back.

CHAPTER 3

METHODOLOGY

3.1 Introduction

Aim of this chapter is to introduce the basic terminologies, theories. Later data collection procedure and methodology is described in details.

3.2 Research Subject and Instrumentation

3.2.1 Correlation

In statistics, correlation is way that helps to measure how two variables are linearly related. Values of the correlation can be either positive or negative. If the values are positive then it means that if value of independent variable increases then the value of dependent variable also increase with respect to the value of correlation. In case of negative correlation, it works completely opposite.

3.2.2 Univariate feature selection

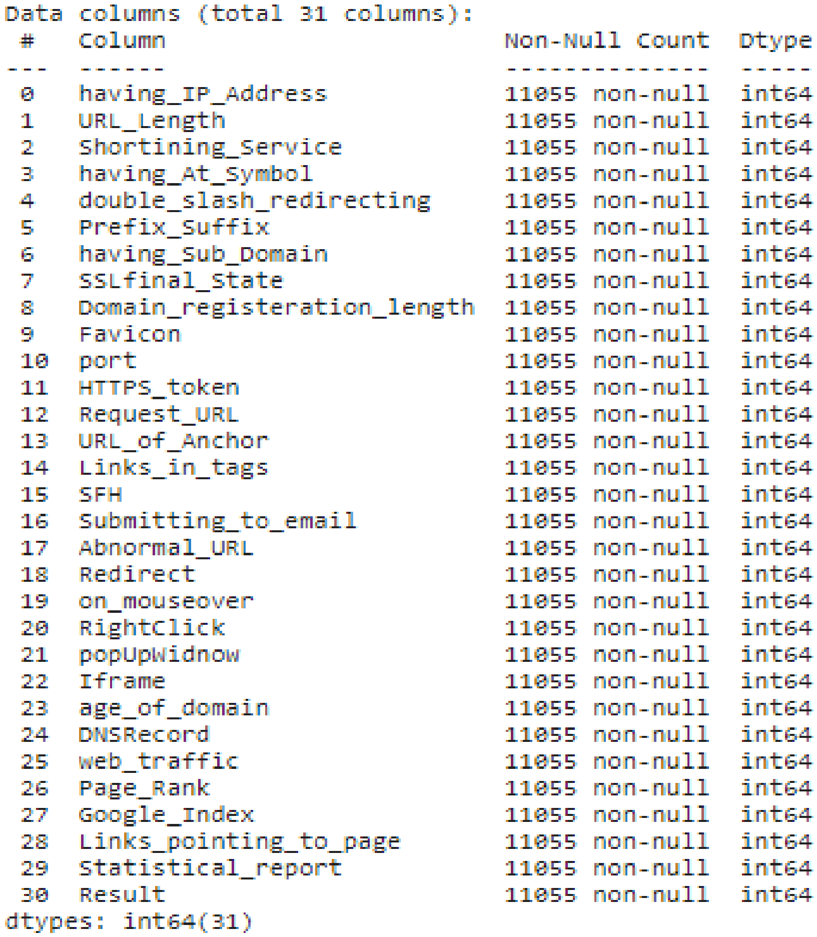
Univariate feature selection examines each feature individually to determine the strength of the relationship of the feature with the response variable.

3.2.3 Feature importance

Feature importance refers to techniques that assign a score to input features based on how useful they are at predicting a target variable.

3.3 Data Collection Procedure

Data was collected from UCI machine learning data repository [10]. After checking whether there is any null values or any duplicate data it was found that there are 31 attributes and 11055 instances in the dataset. List of the features are presented in figure 3.1.



```
Data columns (total 31 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   having_IP_Address            11055 non-null  int64
 1   URL_Length                   11055 non-null  int64
 2   Shortining_Service           11055 non-null  int64
 3   having_At_Symbol             11055 non-null  int64
 4   double_slash_redirecting     11055 non-null  int64
 5   Prefix_Suffix                11055 non-null  int64
 6   having_Sub_Domain            11055 non-null  int64
 7   SSLfinal_State               11055 non-null  int64
 8   Domain_registeration_length  11055 non-null  int64
 9   Favicon                      11055 non-null  int64
 10  port                         11055 non-null  int64
 11  HTTPS_token                  11055 non-null  int64
 12  Request_URL                  11055 non-null  int64
 13  URL_of_Anchor                11055 non-null  int64
 14  Links_in_tags                11055 non-null  int64
 15  SFH                          11055 non-null  int64
 16  Submitting_to_email          11055 non-null  int64
 17  Abnormal_URL                 11055 non-null  int64
 18  Redirect                     11055 non-null  int64
 19  on_mouseover                 11055 non-null  int64
 20  RightClick                   11055 non-null  int64
 21  popUpWidnow                  11055 non-null  int64
 22  Iframe                       11055 non-null  int64
 23  age_of_domain                11055 non-null  int64
 24  DNSRecord                    11055 non-null  int64
 25  web_traffic                  11055 non-null  int64
 26  Page_Rank                    11055 non-null  int64
 27  Google_Index                 11055 non-null  int64
 28  Links_pointing_to_page       11055 non-null  int64
 29  Statistical_report           11055 non-null  int64
 30  Result                       11055 non-null  int64
dtypes: int64(31)
```

Figure 3.1: List of features of the dataset

3.4 Research Methodology

Research methodology for this research shown in figure 3.1 includes 5 steps. They are as follows:

i) Data preprocessing
ii) Data analysis
iii) Feature selection
iv) Logic building
v) System evaluation

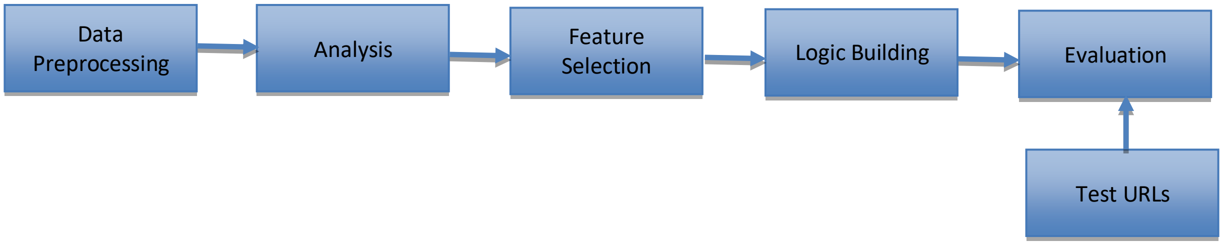


Figure 3.2: Research methodology

i) Data preprocessing: Data preprocessing includes finding the null or missing values in our dataset.

ii) Data analysis: In this step various exploratory analysis will be applied on the dataset to identify correlation of the features.

iii) Feature selection: Top 5 features will be selected in this stage.

iv) Logic building: Based on the selected features logic will be developed in order to detect the phishing websites.

v) System evaluation: Test URLs will be given input to test where our built system can identify the right sites or not.

3.5 Implementation Requirements

In order to conduct the research, we have used Google Colab and Python language. Among the libraries used for data preprocessing, visualization, correlation and univariate analysis are listed below:

i) Pandas

ii) NumPy

iii) Matplotlib

iv) Seaborn

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental setup

In order to start the experiment first have to preprocess the data. In order to preprocess our dataset, we have used Padas and NumPy libraries. Outcome of the execution of those are presented in

1. Correlation (negative co-relation)
2. Univariate Feature selection (Chi-Square method)
3. Feature importance (Tree based classifier)
4. Selection of features (common features of three tests)
5. Phishing Site Detection System development
6. Experiment
7. Accuracy

4.2 Experimental result and analysis

4.2.1 Correlation

Correlation is way that helps to measure how two variables are linearly related. In order to visualize the correlation between any variable with the class a heatmap was generated which is shown in figure 4.1. In the figure it is observed that the deeper the color is the more they are positively correlated and the lighter they are the more they are negatively correlated. In our case we are looking for negative correlation as in our dataset phishing sites are represented by -1. After observing all the correlation top ten most highly correlated variables are then figured out and later represented in figure 4.2.

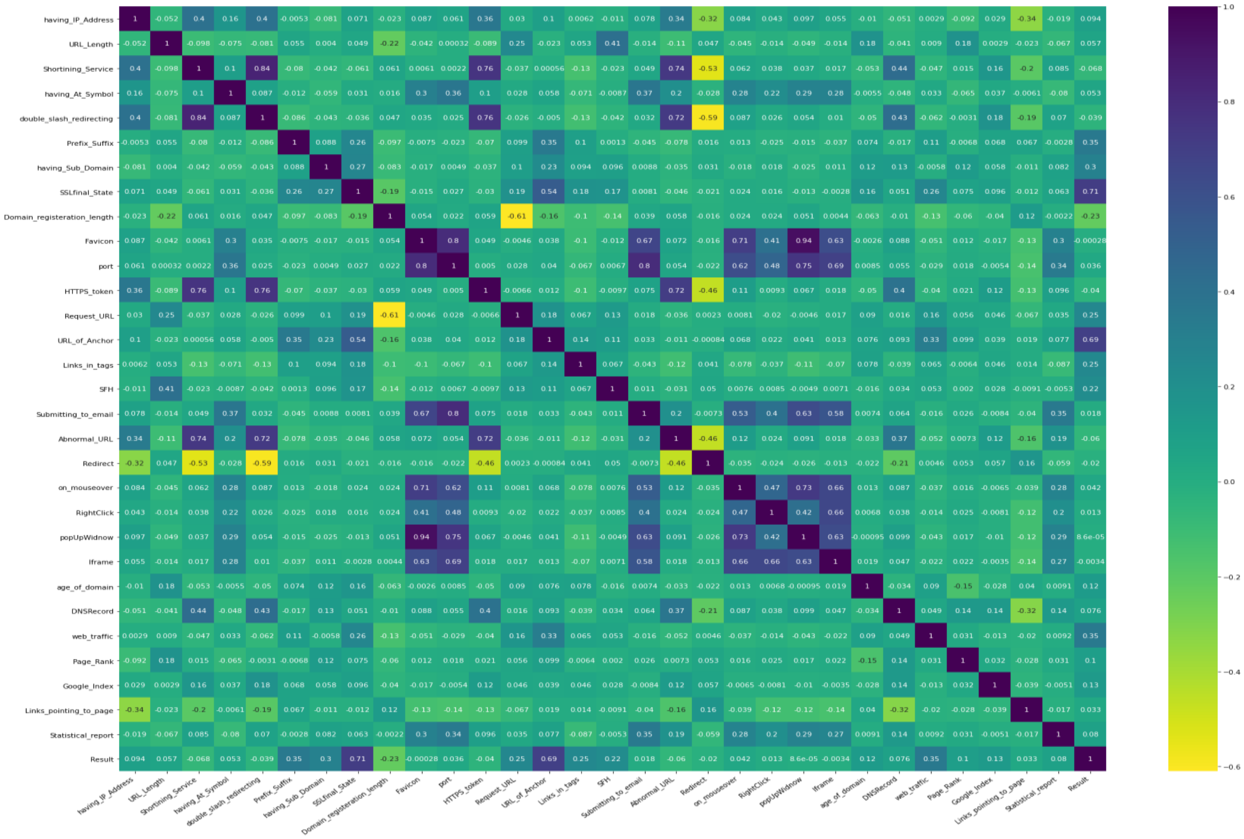


Figure 4.1: Heatmap of correlation of variable between class variable

From the correlation measurement top ten most highly correlated features are shown in figure 4.2.

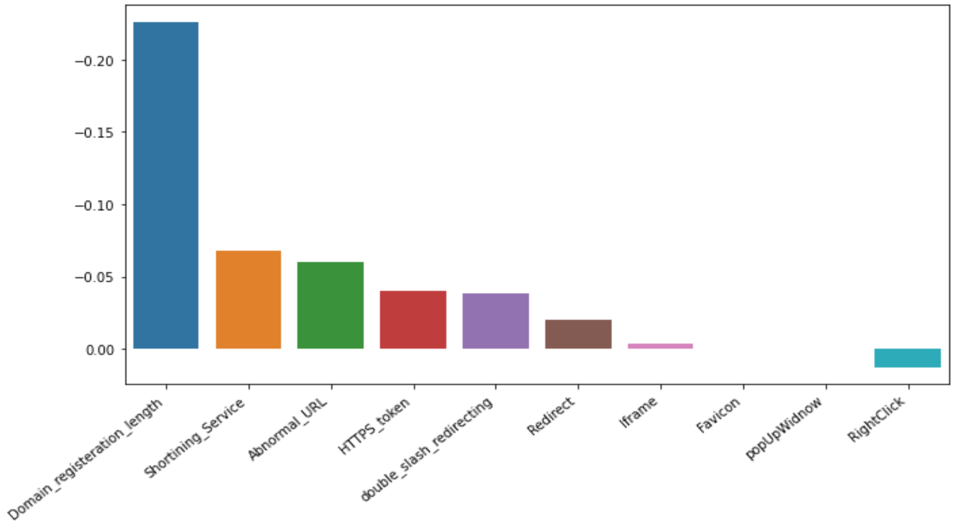


Figure 4.2: Top ten most highly correlated features

4.2.2 Univariate Feature Selection

Univariate feature selection examines each feature individually to determine the strength of the relationship of the feature with the response variable. After examining our dataset using univariate feature selection, top ten features are presented in figure 4.3.

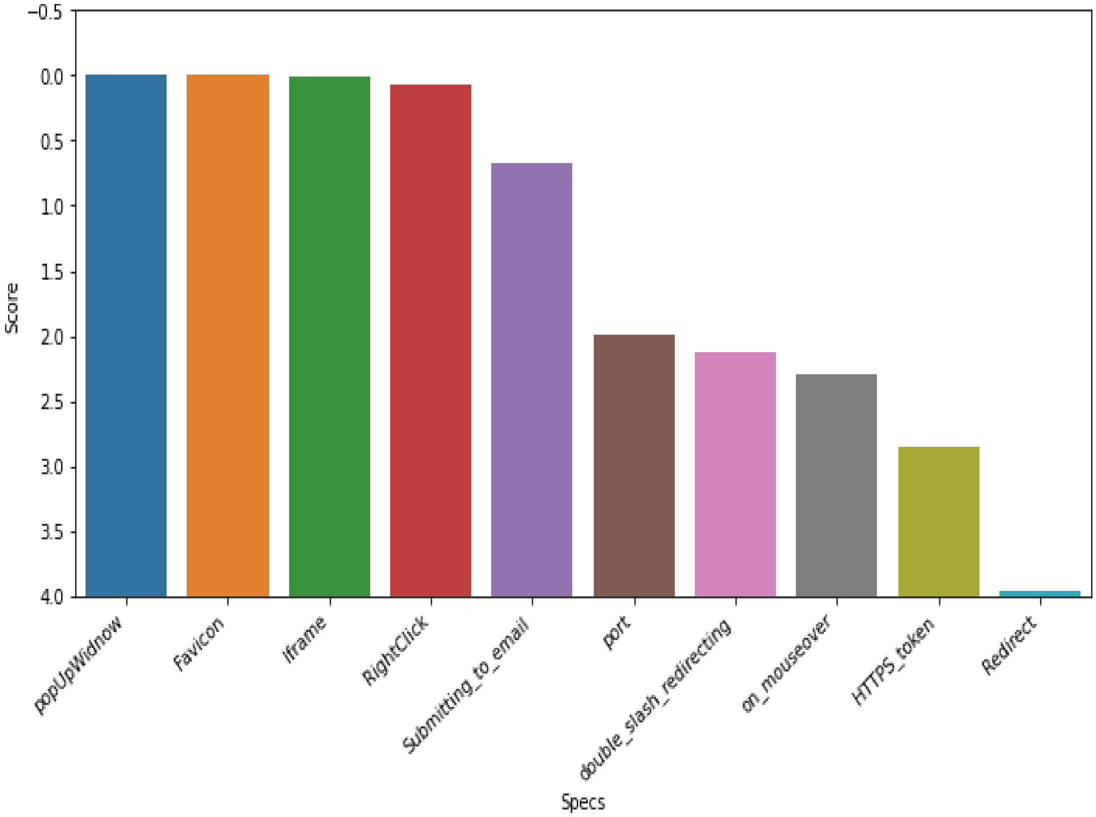


Figure 4.3: Top ten features selected by univariate feature selection method

4.2.3 Feature Importance

Feature importance refers to techniques that assign a score to input features based on how useful they are at predicting a target variable. After testing each feature top ten most important features are selected which are shown in figure 4.4.

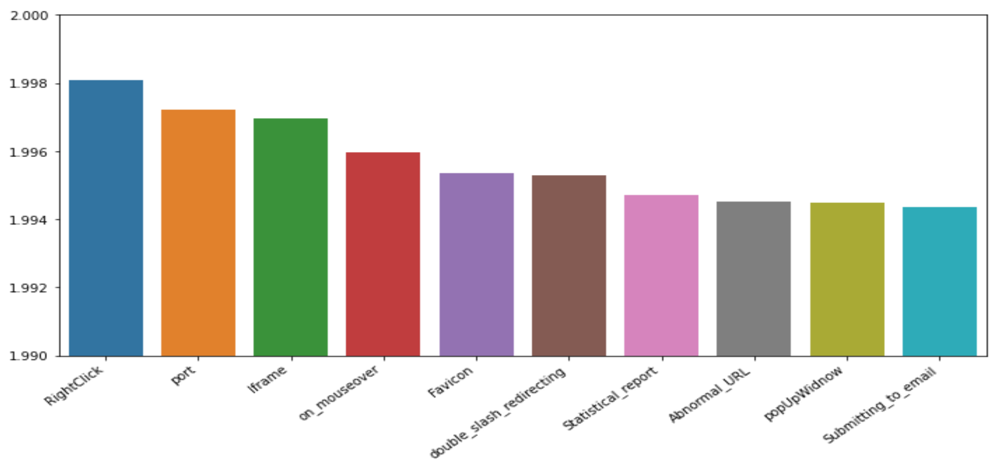


Figure 4.4: top ten selected features by feature importance method

4.2.4 Feature selection

After completing all the above test features were selected. In order to select the features, it was considered that selected feature should be identified by at least two feature selection method. Based on this condition the following features were selected:

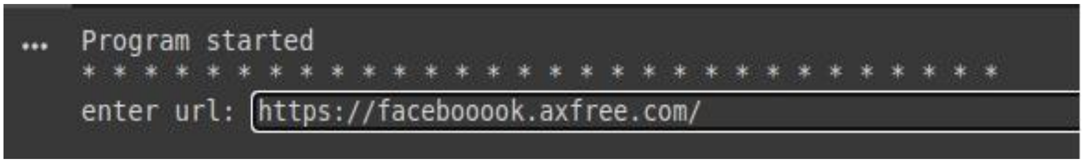
Table 4.1: List of selected features

Feature	Correlation	Univariate	Feature importance
double_slash_redirecting	Y	Y	Y
Iframe	Y	Y	Y
Favicon	Y	Y	Y
RightClick	Y	Y	Y
Shortining_Service	Y	N	Y
Abnormal_URL	Y	Y	N
HTTPS_token	Y	Y	N
popUpWidnow	Y	Y	Y
Submitting_to_email	N	Y	Y
Port	N	Y	Y
on_mouseover	N	Y	Y
Statistical_report	N	Y	Y

4.2.5 Phishing Site Detection System

Though the target of the research was to explore the most significant features to identify a phishing site, we tried to implement our finding to develop a system that may help to identify a phishing site. As a result, we built a system that will take input of a website and in return our system will give feedback whether that site is a phishing site or not.

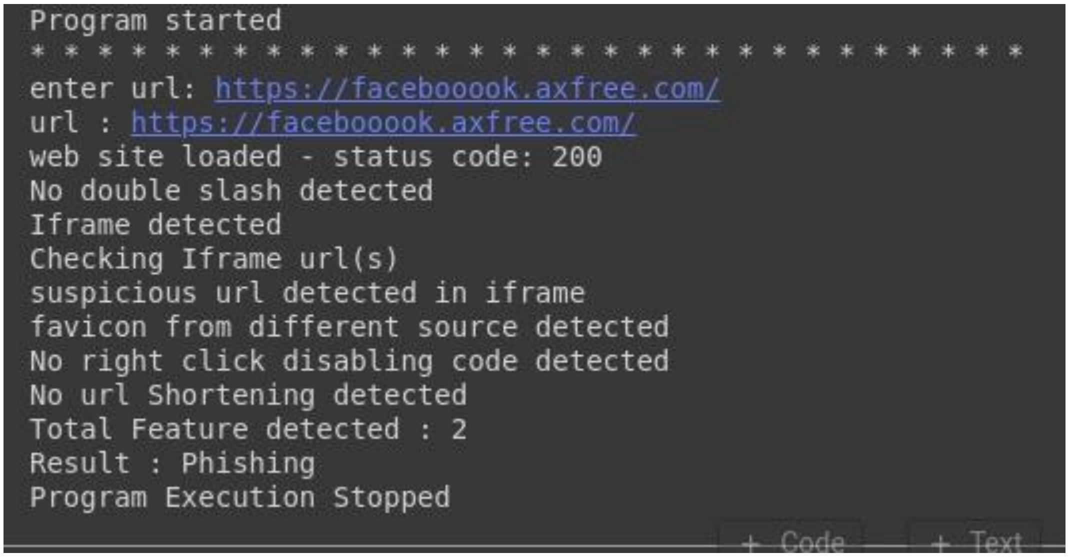
As its very difficult to find a live phishing site manually, we could not test the system with the help of a live phishing site. However, test outcome on a non-live phishing site proved that it can identify it as a phishing site. Figure 4.5 shows that an URL is given input to test whether it is a phishing site or not.



```
...  Program started
     * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
     enter url: https://faceboook.axfree.com/
```

Figure 4.5: Testing a site

Figure 4.6 shows the outcome of the testing.



```
Program started
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
enter url: https://faceboook.axfree.com/
url : https://faceboook.axfree.com/
web site loaded - status code: 200
No double slash detected
Iframe detected
Checking Iframe url(s)
suspicious url detected in iframe
favicon from different source detected
No right click disabling code detected
No url Shortening detected
Total Feature detected : 2
Result : Phishing
Program Execution Stopped
```

Figure 4.6: Outcome of the testing

4.3 Discussion

In order to get the most significant features to identify a phishing site we have applied three feature selection methods. They are correlation, univariate feature selection and feature importance method. After applying all these methods, the feature selected by at least 2 methods were finally selected to identify a phishing website. The finally selected features are:

- Iframe
- Favicon
- RightClick
- Shortining_Service
- Abnormal_URL
- HTTPS_token
- popUpWidnow
- Submitting_to_email
- Port
- on_mouseover
- Statistical_report

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

Phishing is not only a common cybercrime but a top ranked cybercrime in recent times. Every person who is in touch of internet and browsing in the web may become a victim of this crime anytime. Therefore, research in this aspect is very essential. It will help nonprofessional internet users to identify the phishing websites easily.

5.2 Impact on Environment

Our research may not have a direct connection with environment but it certainly will have some on humans. Young and non-technology expert entrepreneurs will be able save a lot of their customers and money by the outcome of this research. As a result, live will live a better and less stressed life.

5.3 Ethical Aspects

Our research project provides a standard system that saves the members of our society from being deceived. Moreover, we have considered every point of research ethics in the project. None of the procedures we have followed during the research have violated any moral values of our society. After the deployment of the system, it can be used without harming any ethics or morals as it simply helps to prevent a cybercrime using technology. Thus, it can be stated that our research project is an ethical and feasible system developed with the help of modern technology.

5.4 Sustainability Plan

The system we developed is moderately sustainable as per our opinion. As its performance depends on how accurately the features are selected. Besides criminals changes their phishing techniques every now and then. Therefore, to get a reliable and trustworthy phishing website detecting system it is very important to update the dataset every now and then. Otherwise, it will be difficult to get a reliable output from this research.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the study

Our project is a research-based project which aims to identify the most important features of phishing websites and to implement those to detect phishing websites. In order to achieve the result exploratory analyses were conducted first to distinguish most important features of the dataset. Later these features were used to build a system that will be able identify whether any given url belongs to phishing website or not.

6.2 Conclusion

At the end of the discussion we can conclude that the final outcome of our research is a system that will enable users to identify phishing websites. As a result, it is expected that our system will help to decrease one of the common cybercrimes of the world if used by the users. Therefore, it will have a great impact both on our economical and social life.

6.3 Implication for further study

The key points that we are keeping in mind for the future improvements of our project are stated below: -

- Expanding the dataset: More updated datasets will definitely help to figure out new and more reliable features for detecting phishing websites.
- Other updated algorithms and machine learning techniques can be applied for study in order to understand whether any other features can help to identify phishing websites better or not.

References

1. Roopak, S., & Thomas, T. “A Novel Phishing Page Detection Mechanism Using HTML Source Code Comparison and Cosine Similarity.” In Advances in Computing and Communications (ICACC), 2014 Fourth International Conference on (pp. 167-170). IEEE.

2. Aburrous, M., Hossain, M. A., Thabatah, F., &Dahal, K. “Intelligent phishing website detection system using fuzzy techniques.” In Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on (pp. 1-6). IEEE.

3. Geng, G. G., Lee, X. D., Wang, W., & Tseng, S. S.). "Favicon-a clue to phishing sites detection." In eCrime Researchers Summit (eCRS), 2013 (pp. 1-10). IEEE.

4. Nguyen, L. A. T., To, B. L., Nguyen, H. K., & Nguyen, M. H. "An efficient approach for phishing detection using single-layer neural network." In Advanced Technologies for Communications (ATC), 2014 International Conference on (pp. 435-440). IEEE.

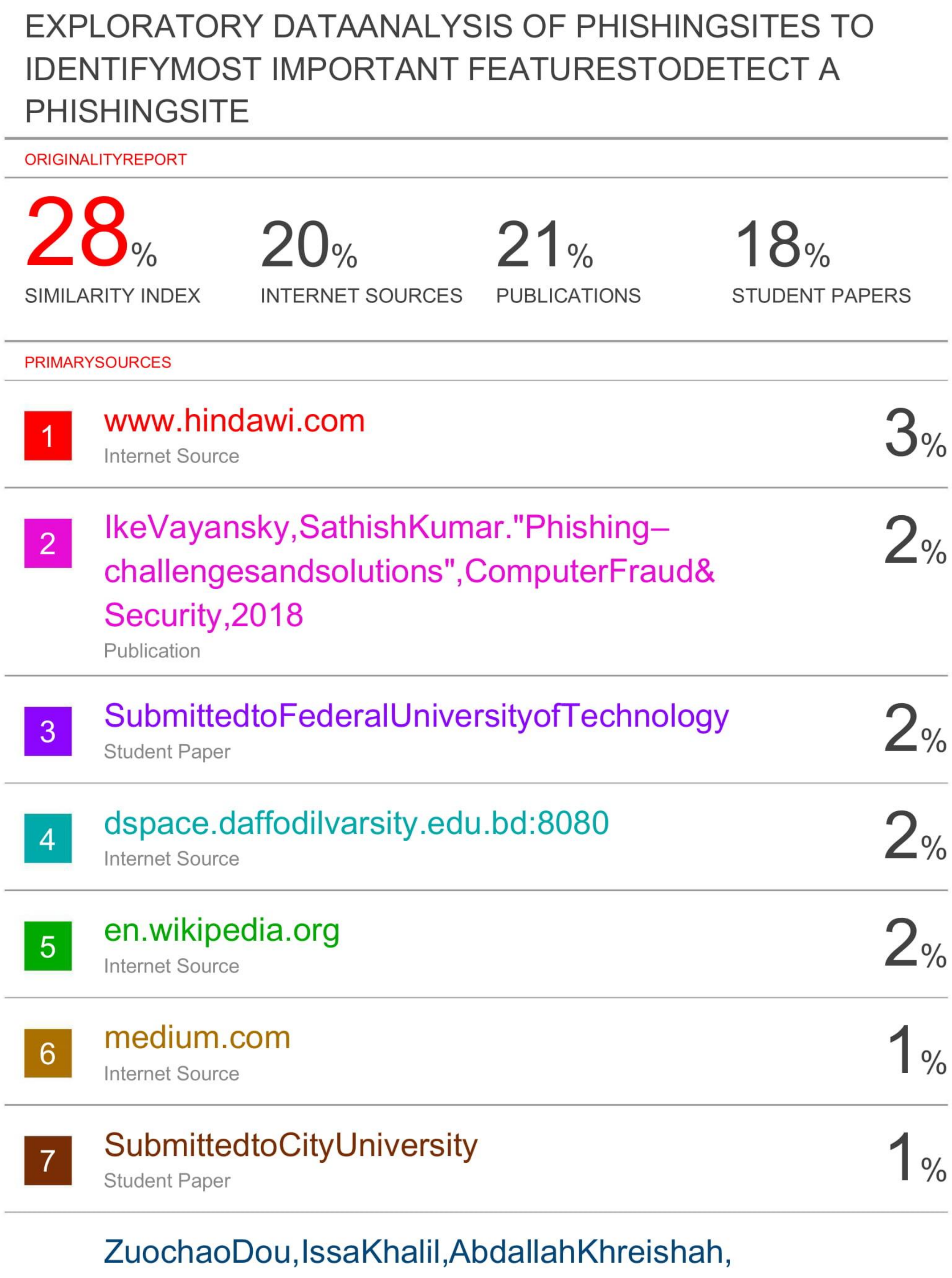
5. Pan, Y., & Ding, X. "Anomaly based web phishing page detection." In Computer Security Applications Conference, 2006. ACSAC'06. 22nd Annual (pp. 381-392). IEEE.

6. Afroz, S., &Greenstadt, R. (2011, September). Phishzoo: Detecting phishing websites by looking at them. In 2011 IEEE fifth international conference on semantic computing (pp. 368-375). IEEE.

7. Parekh, S., Parikh, D., Kotak, S., &Sankhe, S. (2018, April). A new method for detection of phishing websites: URL detection. In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT) (pp. 949-952). IEEE.
8. Cyber Security threats rankings. Available at: https://securityboulevard.com/2020/08/top-5-cybersecurity-threats-2020-what-ranks-alongside-ransomware-and-office-suite-account-hijacking/ [Last access on 6 Octber, 2020]

9. Cyber Security threats. Available at: https://www.thesslstore.com/blog/the-top-9-cyber-security-threats-that-will-ruin-your-day/ [Last access on 6 Octber, 2020]

10. Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA:



EXPLORATORY DATAANALYSIS OF PHISHINGSITES TO IDENTIFYMOST IMPORTANT FEATURESTODETECT A PHISHINGSITE

ORIGINALITYREPORT

28%	20%	21%	18%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARYSOURCES

1	www.hindawi.com Internet Source	3%
2	IkeVayansky,SathishKumar."Phishing–challengesandsolutions",ComputerFraud& Security,2018 Publication	2%
3	SubmittedtoFederalUniversityofTechnology Student Paper	2%
4	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
5	en.wikipedia.org Internet Source	2%
6	medium.com Internet Source	1%
7	SubmittedtoCityUniversity Student Paper	1%
	ZuochaoDou,IssaKhalil,AbdallahKhreishah,	