iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: V Month of publication: May 2020

DOI: http://doi.org/10.22214/ijraset.2020.5378

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

A Novel Machine Learning & NLP based approach for Analysing Phishing Attack

Dr. Savita Sangam[1], Rajshree Shejwal [2], Gouri Shelar [3]
[1, 2, 3] Dept. of Information Technology Engineering, SSJCOE, Maharashtra, India

Abstract: Phishing refers to a sophisticated social networking technique used to target sensitive victims. Now the days of people who use the internet to do many things like shopping, advertising, shipping etc. Phishing is a major attack on the website, people facing in their day to day life. Phishers uses custom-looking websites similar to those of original websites. As technology continues to grow, phishing scams are beginning to develop rapidly and this needs to be prevented by using phishing techniques. Machine learning is a powerful tool used to combat phishing attacks. This paper looks at the materials used to derive detection and detection methods using machine learning and NLP

Keywords: Phishing, Detection, Machine Learning, NLP

I. INTRODUCTION

In our society, the security of confidential information is a matter that concerns everyone. Phishing is a type of attack on social engineers that focuses on obtaining sensitive information in disguise as an organization worthy of trust. Electronic communications, such as email or text messaging are common platforms for sending phishing attacks. Attackers are often disguised as popular social websites, banks, and executives from the IT department or online shopping websites. Most methods check the URLs that contain the message. Our approach performs a non-linear analysis of the text transmitted by the attacker to verify the validity of each paragraph. Sentence is considered dangerous if it requires sensitive information or dictates the performance of an action that would disclose personal information. The roles of each word in a sentence, the way we use it determines whether a sentence is a question or a command. Much research has been ongoing to prevent phishing attacks by various communities around the world. Phishing attacks can be prevented by detecting websites and notifying users to detect phishing websites. Machine learning algorithms have been one of the most effective techniques for detecting phishing websites. In this study, various ways to access phishing websites were discussed.

II. LITERATURE SURVEY

A. TianruiPeng & et. All Data Hacking Attacks sensitive data using natural language processing and machine learning. We are introducing a way to detect cybercrime attacks. Our approach relies on text analysis, rather than the metadata that may be associated with emails. As a result, our method is effective for obtaining phishing emails containing clean text. Our results on cybercrime emails show the most advanced recollections that indicate that semantic information is a strong indicator of communication engineering. Electronic communications, such as email or text messaging are common platforms for sending phishing attacks. Phishing has been shown as an effective attack throughout the years, fooling many people. Attackers are often disguised as popular social websites, banks, and executives from the IT department or online shopping websites. These emails may urge users to click links to perform malware downloads, or enter personal information on a malicious website that has the same legal appearance. Many ways to get automated email automation are based on email metadata, data associated with emails not related to the semantic meaning of the text. Most methods check the URLs that contain the message.

B. Phish Storm: Identifying Phishing Analytics by Streaming Analytics Samuel Marshal & et all Despite the rise of prevention methods, phishing is currently a significant threat because the primary accounting used is based on active URL testing. This method does not work well due to the short time of designing Web sites, making recent methods dependent on the use of real-time URLs or techniques for obtaining relevant criminal information. In this paper we introduce Phish Storm, a phishing scheme that can process real-time any URL to identify phishing sites. Phish Storm can interact with any email server or HTTP proxy. We argue that phishing URLs generally have little to do between the part of the URL to be registered (the base domain) and the remainder of the URL (the top-level domain, method, query). We show in this paper that experimental evidence supports this observation and can be used to identify phishing sites. For this purpose, we define a new concept, related to URL combinations and test it using features extracted from URL names based on query data from Google and Yahoo search engines. These features are used for machine learning classification to retrieve phishing URLs from sensitive data.

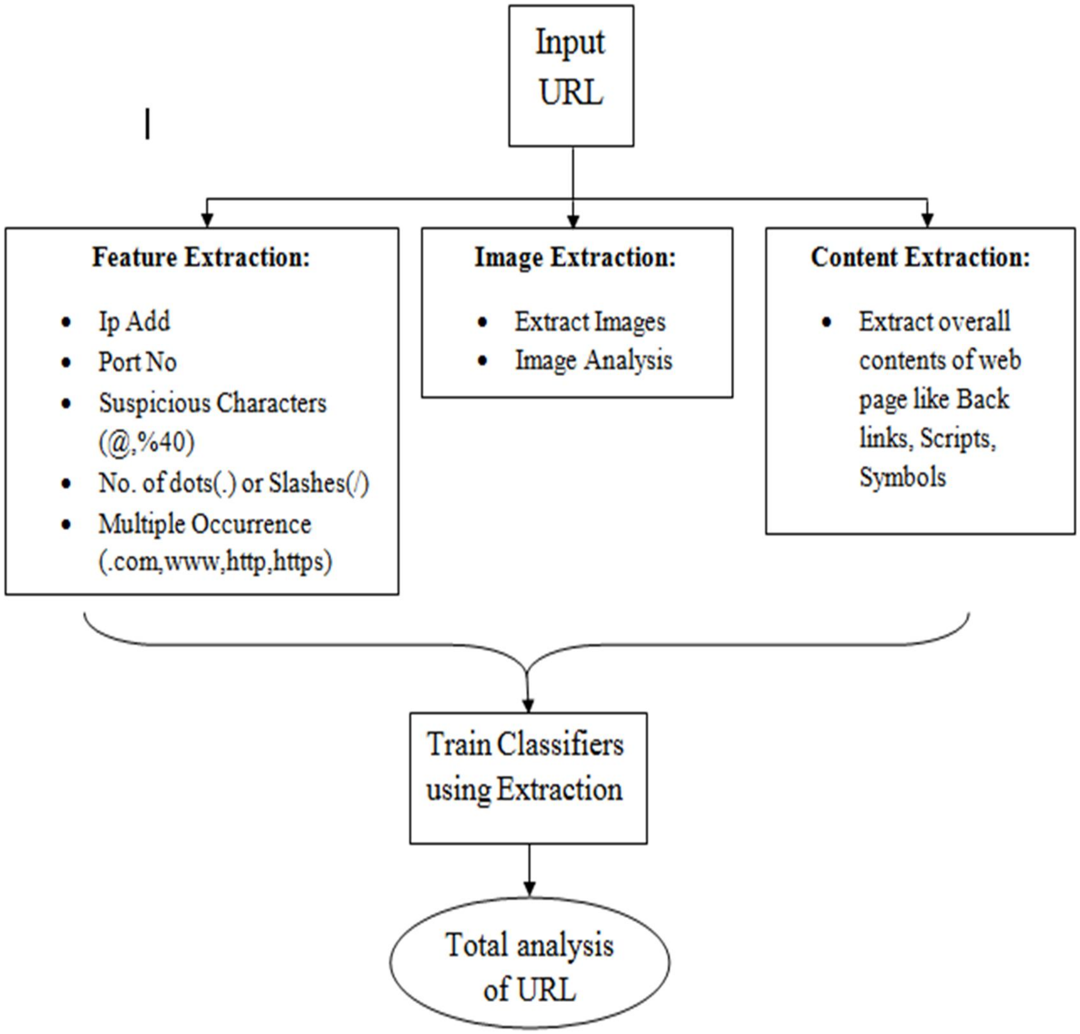
C. Phishing URL discovery: web mining learning method. B.E. Sananse & et all 2016 the identification of malicious queries and commands depends on the presence of a list of topics in the verb-direct object in pairs where their presence of query command is intended to be malicious. To generate a list of adults in an article we use machine learning, build a Naive Bayes data-driven class that is distributed across multiple countries, and is used to classify text. We used the Multinomial () function from the Scikit-learn Python library that uses this algorithm. This algorithm generates each prediction label (the verb of a specific object), and produces a confidence note in the prediction. Range of confidence scores is 0 to 1, with a validation verification number. We used 1000 phishing emails from Nazi phishing emails. We tested our results on all 5014 emails in the Nazi phishing email. As well as about 5000 phishing emails on Enron Corpus. Before applying for machine learning, we used Stanford typed water parser to extract all (verb-direct object) from all expressions by identifying “nsubjpass” and “dobj” based on the findings of each sentence.

III. IMPLEMENTATION

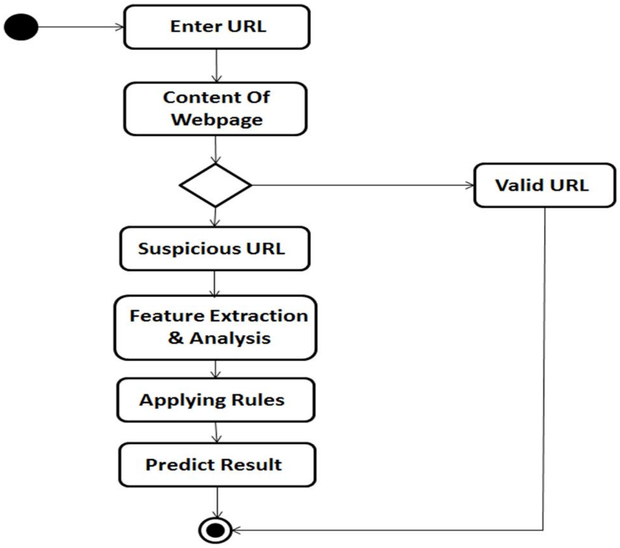
Phishing attacks are one of the most common and least defended security threats today. We present an approach which uses natural language processing techniques to analyze text and detect inappropriate statements which are indicative of phishing attacks. Our approach is novel compared to previous work because it focuses on the natural language text contained in the attack, performing semantic analysis of the text to detect malicious intent. To demonstrate the effectiveness of our approach, we have evaluated it using a large benchmark set of phishing URL.

A machine learning based anti-phishing system which is based on Uniform Resource Locator (URL) features.

A. We Have Taken Some Features


Input URL
Feature Extraction:
Ip Add
Port No
Suspicious Characters (@,%40)
No. of dots(.) or Slashes(/)
Multiple Occurrence (.com,www,http,https)
Image Extraction:
Extract Images
Image Analysis
Content Extraction:
Extract overall contents of web page like Back links, Scripts, Symbols
Train Classifiers using Extraction
Total analysis of URL

Below is a only shows the flow of data between user and the system. The user enter URL .the feature extraction are applying to URL and check the URL length, number of suspicious character, multiple occurrence of symbols and protocol.



B. *Detecting Accuracy of URL*

1) *IP Address:* Example: “ http://66.135.200.145 ”
2) *Protocol:* Example: http, https, ftp, SSH, Telnet Natural Language Processing (NLP)

a) *Step1:* The input to natural language processing will be a simple stream of Unicode characters (typically UTF-8). Basic processing will be required to convert this character stream into a sequence of lexical items (words, phrases, and syntactic markers) which can then be used to better understand the content.

b) *Step2:* Decide on Micro Understanding It is used for: clustering, categorization, similarity, topic analysis, word clouds and summarization.

C. *Numbers of Dots and Slashes*

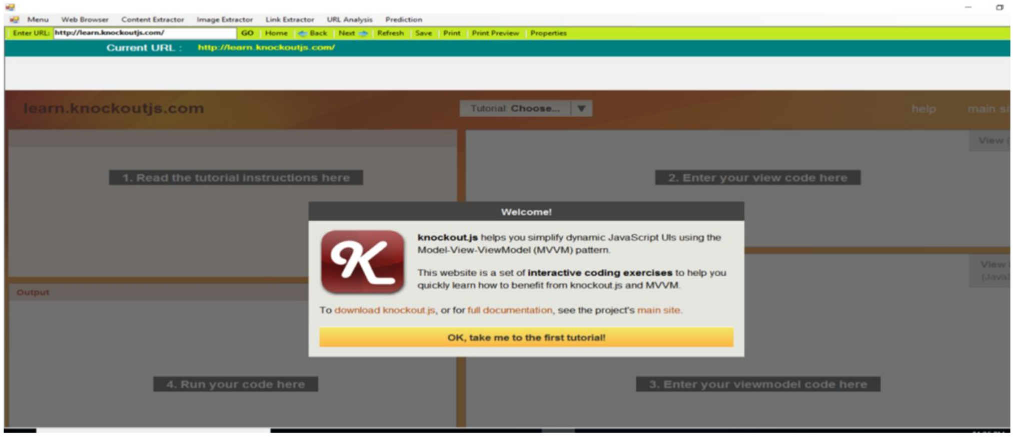
Example: *https://www.**idbi**.com*

https://www.ldbi.com/.in

1) Multiple Occurrence(.com ,https ,http):
2) Example: *http://www.google.com/url?q=http://www.badsite.com*

IV. RESULT

1) *User Browser:* In this diagram user enter the URL and perform their activity but this time more chances of fake URLs and redirected links are attached in valid URL, so our project help us to user to detect the valid URLs



Enter URL: http://learn.knockoutjs.com/

Current URL : http://learn.knockoutjs.com/

Welcome!

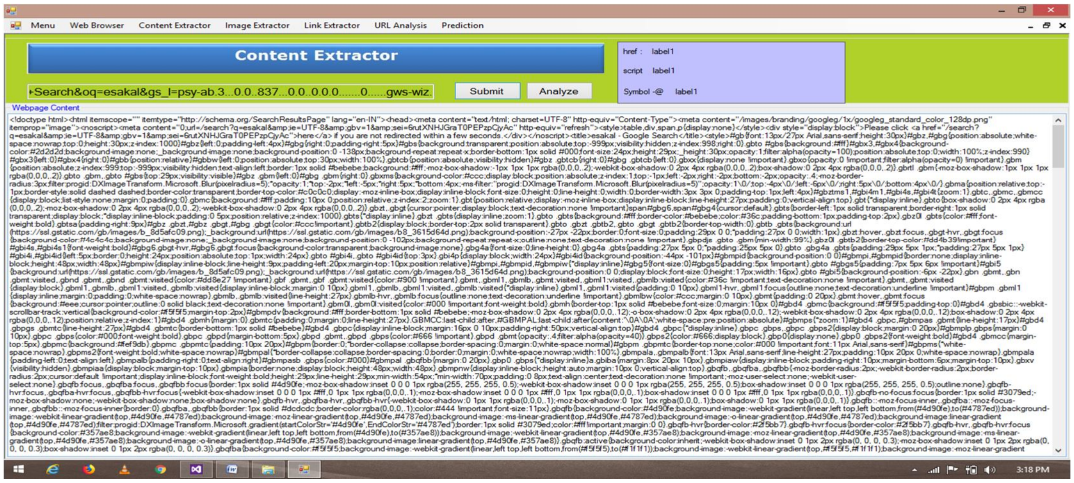
knockout.js helps you simplify dynamic JavaScript UIs using the Model-View-ViewModel (MVVM) pattern.

This website is a set of **interactive coding exercises** to help you quickly learn how to benefit from knockout.js and MVVM.

To download knockout.js, or for full documentation, see the project's main site.

Fig 4.1 User Browser

2) *Content Extractor:* In this diagram current URL are used and all content are collected in currant URL then by using DOM Href , Links, Symbols are separated and showing the count.



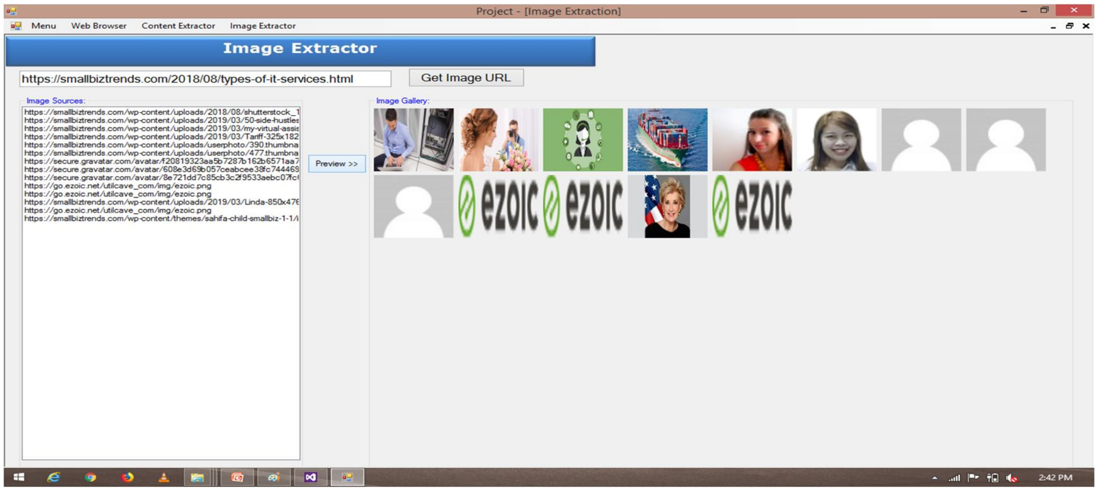
+Search&oq=esakal&gs_l=psy-ab.3...0.0..837...0.0..0.0.0.......0......gws-wiz.

href : label1
script label1
Symbol -@ label1

```
<!doctype html><html itemscope="" itemtype="http://schema.org/SearchResultsPage" lang="en-IN"><head><meta content="text/html; charset=UTF-8" http-equiv="Content-Type"><meta content="/images/branding/googleg/1x/googleg_standard_color_128dp.png"
itemprop="image"><noscript><meta content="0;url=/search?q=esakal&amp;ie=UTF-8&amp;gbv=1&amp;sei=6rutXNHJGraT0PEPzpCjyAc" http-equiv="refresh"><style>table,div,span,p{display:none}</style><div style="display:block">Please click <a href="/search?
q=esakal&amp;ie=UTF-8&amp;gbv=1&amp;sei=6rutXNHJGraT0PEPzpCjyAc">here</a> if you are not redirected within a few seconds.</div></noscript><title>esakal - Google Search</title><style>#gb{font:13px/27px Arial,sans-serif;height:30px}#gbz,#gbg{position:absolute;white-
space:nowrap;top:0;height:30px;z-index:1000}#gbz{left:0;padding-left:4px}#gbg{right:0;padding-right:5px}#gbs{background:transparent;position:absolute;top:-999px;visibility:hidden;z-index:998;right:0}.gbto #gbs{background:#fff}#gbx3,#gbx4{background-
color:#2d2d2d;background-image:none;_background-image:none;background-position:0 -138px;background-repeat:repeat-x;border-bottom:1px solid #000;font-size:24px;height:29px;_height:30px;opacity:1;filter:alpha(opacity=100);position:absolute;top:0;width:100%;z-index:990}
#gbx3{left:0}#gbx4{right:0}#gbb{position:relative}#gbbw{left:0;position:absolute;top:30px;width:100%}.gbtcb{position:absolute;visibility:hidden}#gbz .gbtcb{right:0}#gbg .gbtcb{left:0}.gbxx{display:none !important}.gbxo{opacity:0 !important;filter:alpha(opacity=0) !important}.gbm
{position:absolute;z-index:999;top:-999px;visibility:hidden;text-align:left;border:1px solid #bebebe;background:#fff;-moz-box-shadow:-1px 1px 1px rgba(0,0,0,.2);-webkit-box-shadow:0 2px 4px rgba(0,0,0,.2);box-shadow:0 2px 4px rgba(0,0,0,.2)}.gbrtl .gbm{-moz-box-shadow:1px 1px 1px
rgba(0,0,0,.2)}.gbto .gbm,.gbto #gbs{top:29px;visibility:visible}#gbz .gbm{left:0}#gbg .gbm{right:0}.gbxms{background-color:#ccc;display:block;position:absolute;z-index:1;top:-1px;left:-2px;right:-2px;bottom:-2px;opacity:.4;-moz-border-
radius:3px;filter:progid:DXImageTransform.Microsoft.Blur(pixelradius=5);*opacity:1;*top:-2px;*left:-5px;*right:5px;*bottom:4px;-ms-filter:"progid:DXImageTransform.Microsoft.Blur(pixelradius=5)";opacity:1\0/;top:-4px\0/;left:-6px\0/;right:5px\0/;bottom:4px\0/}.gbma{position:relative;top:-
1px;border-style:solid dashed dashed;border-color:transparent;border-top-color:#c0c0c0;display:-moz-inline-box;display:inline-block;font-size:0;height:0;line-height:0;width:0;border-width:3px 3px 0;padding-top:1px;left:4px}#gbztms1,#gbi4m1,#gbi4s,#gbi4t{zoom:1}.gbtc,.gbmc,.gbmcc
{display:block;list-style:none;margin:0;padding:0}.gbmc{background:#fff;padding:10px 0;position:relative;z-index:2;zoom:1}.gbt{position:relative;display:-moz-inline-box;display:inline-block;line-height:27px;padding:0;vertical-align:top}.gbt{*display:inline}.gbto{box-shadow:0 2px 4px rgba
(0,0,0,.2);-moz-box-shadow:0 2px 4px rgba(0,0,0,.2);-webkit-box-shadow:0 2px 4px rgba(0,0,0,.2)}.gbzt,.gbgt{cursor:pointer;display:block;text-decoration:none !important}span#gbg6,span#gbg4{cursor:default}.gbts{border-left:1px solid transparent;border-right:1px solid
transparent;display:block;*display:inline-block;padding:0 5px;position:relative;z-index:1000}.gbts{*display:inline}.gbzt .gbts{display:inline;zoom:1}.gbto .gbts{background:#fff;border-color:#bebebe;color:#36c;padding-bottom:1px;padding-top:2px}.gbz0l .gbts{color:#fff;font-
weight:bold}.gbtsa{padding-right:9px}#gbz .gbzt,#gbz .gbgt,#gbg .gbgt{color:#ccc!important}.gbtb2{display:block;border-top:2px solid transparent}.gbto .gbzt .gbtb2,.gbto .gbgt .gbtb2{border-top-width:0}.gbtb .gbts{background:url
(https://ssl.gstatic.com/gb/images/b_8d5afc09.png);_background:url(https://ssl.gstatic.com/gb/images/b8_3615d64d.png);background-position:-27px -22px;border:0;font-size:0;padding:29px 0 0;*padding:27px 0 0;width:1px}.gbzt:hover,.gbzt:focus,.gbgt-hvr,.gbgt:focus
{background-color:#4c4c4c;background-image:none;_background-image:none;background-position:0 -102px;background-repeat:repeat-x;outline:none;text-decoration:none !important}.gbpdjs .gbto .gbm{min-width:99%}.gbz0l .gbtb2{border-top-color:#dd4b39!important}
#gbi4s,#gbi4s1{font-weight:bold}#gbg6.gbgt-hvr,#gbg6.gbgt:focus{background-color:transparent;background-image:none}.gbg4a{font-size:0;line-height:0}.gbg4a .gbts{padding:27px 5px 0;*padding:25px 5px 0}.gbto .gbg4a .gbts{padding:29px 5px 1px;*padding:27px 5px 1px}
#gbi4i,#gbi4id{left:5px;border:0;height:24px;position:absolute;top:1px;width:24px}.gbto #gbi4i,.gbto #gbi4id{top:3px}.gbi4p{display:block;width:24px}#gbi4id{background-position:-44px -101px}#gbmpid{background-position:0 0}#gbmpi,#gbmpid{border:none;display:inline-
block;height:48px;width:48px}#gbmpiw{display:inline-block;line-height:9px;padding-left:20px;margin-top:10px;position:relative}#gbmpi,#gbmpid,#gbmpiw{*display:inline}#gbg5{font-size:0}#gbgs5{padding:5px !important}.gbto #gbgs5{padding:7px 5px 6px !important}#gbi5
{background:url(https://ssl.gstatic.com/gb/images/b_8d5afc09.png);_background:url(https://ssl.gstatic.com/gb/images/b8_3615d64d.png);background-position:0 0;display:block;font-size:0;height:17px;width:16px}.gbto #gbi5{background-position:-6px -22px}.gbn .gbmt,.gbn
.gbmt:visited,.gbnd .gbmt,.gbnd .gbmt:visited{color:#dd8e27 !important}.gbf .gbmt,.gbf .gbmt:visited{color:#900 !important}.gbmt,.gbml1,.gbmlb,.gbmt:visited,.gbml1:visited,.gbmlb:visited{color:#36c !important;text-decoration:none !important}.gbmt,.gbmt:visited
{display:block}.gbml1,.gbmlb,.gbml1:visited,.gbmlb:visited{display:inline-block;margin:0 10px}.gbml1,.gbmlb,.gbml1:visited,.gbmlb:visited{*display:inline}.gbml1,.gbml1:visited{padding:0 10px}.gbml1-hvr,.gbml1:focus{outline:none;text-decoration:underline !important}#gbpm .gbml1
{display:inline;margin:0;padding:0;white-space:nowrap}.gbmlb,.gbmlb:visited{line-height:27px}.gbmlb-hvr,.gbmlb:focus{outline:none;text-decoration:underline !important}.gbmlbw{color:#ccc;margin:0 10px}.gbmt{padding:0 20px}.gbmt:hover,.gbmt:focus
{background:#eee;cursor:pointer;outline:0 solid black;text-decoration:none !important}.gbm0l,.gbm0l:visited{color:#000 !important;font-weight:bold}.gbmh{border-top:1px solid #bebebe;font-size:0;margin:10px 0}#gbd4 .gbmc{background:#f5f5f5;padding-top:0}#gbd4 .gbsbic::-webkit-
scrollbar-track:vertical{background-color:#f5f5f5;margin-top:2px}#gbmpdv{background:#fff;border-bottom:1px solid #bebebe;-moz-box-shadow:0 2px 4px rgba(0,0,0,.12);-o-box-shadow:0 2px 4px rgba(0,0,0,.12);-webkit-box-shadow:0 2px 4px rgba(0,0,0,.12);box-shadow:0 2px 4px
rgba(0,0,0,.12);position:relative;z-index:1}#gbd4 .gbmh{margin:0}.gbmtc{padding:0;margin:0;line-height:27px}.GBMCC:last-child:after,#GBMPAL:last-child:after{content:'\0A\0A';white-space:pre;position:absolute}#gbmps{*zoom:1}#gbd4 .gbpc,#gbmpas .gbmt{line-height:17px}#gbd4
.gbpgs .gbmtc{line-height:27px}#gbd4 .gbmtc{border-bottom:1px solid #bebebe}#gbd4 .gbpc{display:inline-block;margin:16px 0 10px;padding-right:50px;vertical-align:top}#gbd4 .gbpc{*display:inline}.gbpc .gbps,.gbpc .gbps2{display:block;margin:0 20px}#gbmplp.gbps{margin:0
10px}.gbpc .gbps{color:#000;font-weight:bold}.gbpc .gbpd{margin-bottom:5px}.gbpd .gbmt,.gbpd .gbps{color:#666 !important}.gbpd .gbmt{opacity:.4;filter:alpha(opacity=40)}.gbps2{color:#666;display:block}.gbp0{display:none}.gbp0 .gbps2{font-weight:bold}#gbd4 .gbmcc{margin-
top:5px}.gbpmc{background:#fef9db}.gbpmc .gbpmtc{padding:10px 20px}#gbpm{border:0;*border-collapse:collapse;border-spacing:0;margin:0;white-space:normal}#gbpm .gbpmtc{border-top:none;color:#000 !important;font:11px Arial,sans-serif}#gbpms{*white-
space:nowrap}.gbpms2{font-weight:bold;white-space:nowrap}#gbmpal{*border-collapse:collapse;border-spacing:0;border:0;margin:0;white-space:nowrap;width:100%}.gbmpala,.gbmpalb{font:13px Arial,sans-serif;line-height:27px;padding:10px 20px 0;white-space:nowrap}.gbmpala
{padding-left:0;text-align:left}.gbmpalb{padding-right:0;text-align:right}#gbmpasb .gbps{color:#000}#gbmpal .gbqfbb{margin:0 20px}.gbp0 .gbps{*display:inline}a.gbiba{margin:8px 20px 10px}.gbmpiaw{display:inline-block;padding-right:10px;margin-bottom:6px;margin-top:10px}.gbxv
{visibility:hidden}.gbmpiaa{display:block;margin-top:10px}.gbmpia{border:none;display:block;height:48px;width:48px}.gbmpnw{display:inline-block;height:auto;margin:10px 0;vertical-align:top}.gbqfb,.gbqfba,.gbqfbb{-moz-border-radius:2px;-webkit-border-radius:2px;border-
radius:2px;cursor:default !important;display:inline-block;font-weight:bold;height:29px;line-height:29px;min-width:54px;*min-width:70px;padding:0 8px;text-align:center;text-decoration:none !important;-moz-user-select:none;-webkit-user-
select:none}.gbqfb:focus,.gbqfba:focus,.gbqfbb:focus{border:1px solid #4d90fe;-moz-box-shadow:inset 0 0 0 1px rgba(255, 255, 255, 0.5);-webkit-box-shadow:inset 0 0 0 1px rgba(255, 255, 255, 0.5);box-shadow:inset 0 0 0 1px rgba(255, 255, 255, 0.5);outline:none}.gbqfb-
hvr:focus,.gbqfba-hvr:focus,.gbqfbb-hvr:focus{-webkit-box-shadow:inset 0 0 0 1px #fff,0 1px 1px rgba(0,0,0,.1);-moz-box-shadow:inset 0 0 0 1px #fff,0 1px 1px rgba(0,0,0,.1);box-shadow:inset 0 0 0 1px #fff,0 1px 1px rgba(0,0,0,.1)}.gbqfb-no-focus:focus{border:1px solid #3079ed;-
moz-box-shadow:none;-webkit-box-shadow:none;box-shadow:none}.gbqfb-hvr,.gbqfba-hvr,.gbqfbb-hvr{-webkit-box-shadow:0 1px 1px rgba(0,0,0,.1);-moz-box-shadow:0 1px 1px rgba(0,0,0,.1);box-shadow:0 1px 1px rgba(0,0,0,.1)}.gbqfb::-moz-focus-inner,.gbqfba::-moz-focus-
inner,.gbqfbb::-moz-focus-inner{border:0}.gbqfba,.gbqfbb{border:1px solid #dcdcdc;border-color:rgba(0,0,0,.1);color:#444 !important;font-size:11px}.gbqfb{background-color:#4d90fe;background-image:-webkit-gradient(linear,left top,left bottom,from(#4d90fe),to(#4787ed));background-
image:-webkit-linear-gradient(top,#4d90fe,#4787ed);background-image:-moz-linear-gradient(top,#4d90fe,#4787ed);background-image:-ms-linear-gradient(top,#4d90fe,#4787ed);background-image:-o-linear-gradient(top,#4d90fe,#4787ed);background-image:linear-gradient
(top,#4d90fe,#4787ed);filter:progid:DXImageTransform.Microsoft.gradient(startColorStr='#4d90fe',EndColorStr='#4787ed');border:1px solid #3079ed;color:#fff!important;margin:0 0}.gbqfb-hvr{border-color:#2f5bb7}.gbqfb-hvr:focus{border-color:#2f5bb7}.gbqfb-hvr,.gbqfb-hvr:focus
{background-color:#357ae8;background-image:-webkit-gradient(linear,left top,left bottom,from(#4d90fe),to(#357ae8));background-image:-webkit-linear-gradient(top,#4d90fe,#357ae8);background-image:-moz-linear-gradient(top,#4d90fe,#357ae8);background-image:-ms-linear-
gradient(top,#4d90fe,#357ae8);background-image:-o-linear-gradient(top,#4d90fe,#357ae8);background-image:linear-gradient(top,#4d90fe,#357ae8)}.gbqfb:active{background-color:inherit;-webkit-box-shadow:inset 0 1px 2px rgba(0, 0, 0, 0.3);-moz-box-shadow:inset 0 1px 2px rgba(0,
0, 0, 0.3);box-shadow:inset 0 1px 2px rgba(0, 0, 0, 0.3)}.gbqfba{background-color:#f5f5f5;background-image:-webkit-gradient(linear,left top,left bottom,from(#f5f5f5),to(#f1f1f1));background-image:-webkit-linear-gradient(top,#f5f5f5,#f1f1f1);background-image:-moz-linear-gradient
```

Fig 4.2 Content Extractor

3) *Image Extractor:* When the user checking content extractor this time all link, symbols, slashes and images are separated by using the DOM tree and all images all display in this Image Extractor.

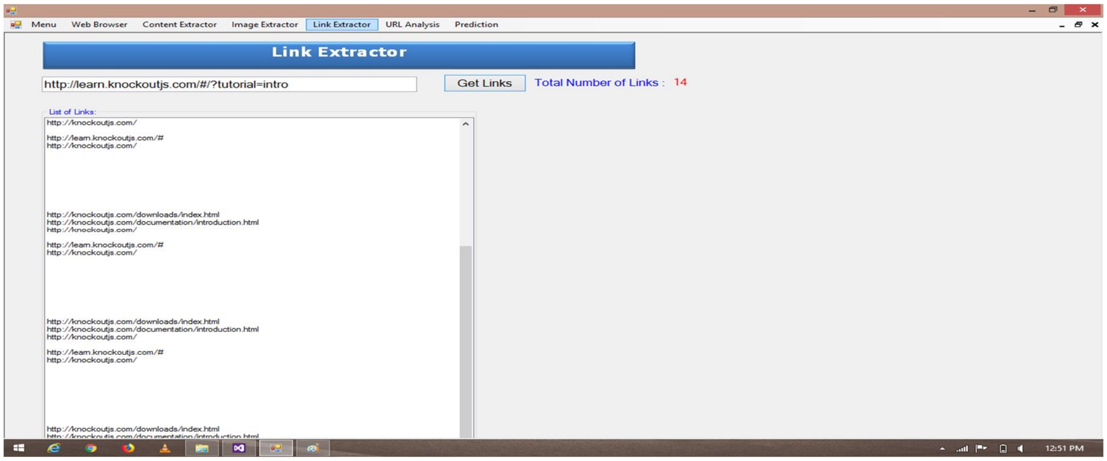


https://smallbiztrends.com/2018/08/types-of-it-services.html

```
https://smallbiztrends.com/wp-content/uploads/2018/08/shutterstock_1
https://smallbiztrends.com/wp-content/uploads/2019/03/50-side-hustles
https://smallbiztrends.com/wp-content/uploads/2019/03/my-virtual-assis
https://smallbiztrends.com/wp-content/uploads/2019/03/Tariff-325x182
https://smallbiztrends.com/wp-content/uploads/userphoto/390.thumbna
https://smallbiztrends.com/wp-content/uploads/userphoto/477.thumbna
https://secure.gravatar.com/avatar/f20819323aa5b7287b162b6571aa7
https://secure.gravatar.com/avatar/608e3d69b057ceabcee38fc744469
https://secure.gravatar.com/avatar/8e721dd7c85cb3c2f9533aebc07fc
https://go.ezoic.net/utilcave_com/img/ezoic.png
https://go.ezoic.net/utilcave_com/img/ezoic.png
https://smallbiztrends.com/wp-content/uploads/2019/03/Linda-850x476
https://go.ezoic.net/utilcave_com/img/ezoic.png
https://smallbiztrends.com/wp-content/themes/sahifa-child-smallbiz-1-1/i
```

Fig 4.3 Image Extractor

4) *Link Extractor:* In this diagram all links are extracted and showing the total no of links contain this URL



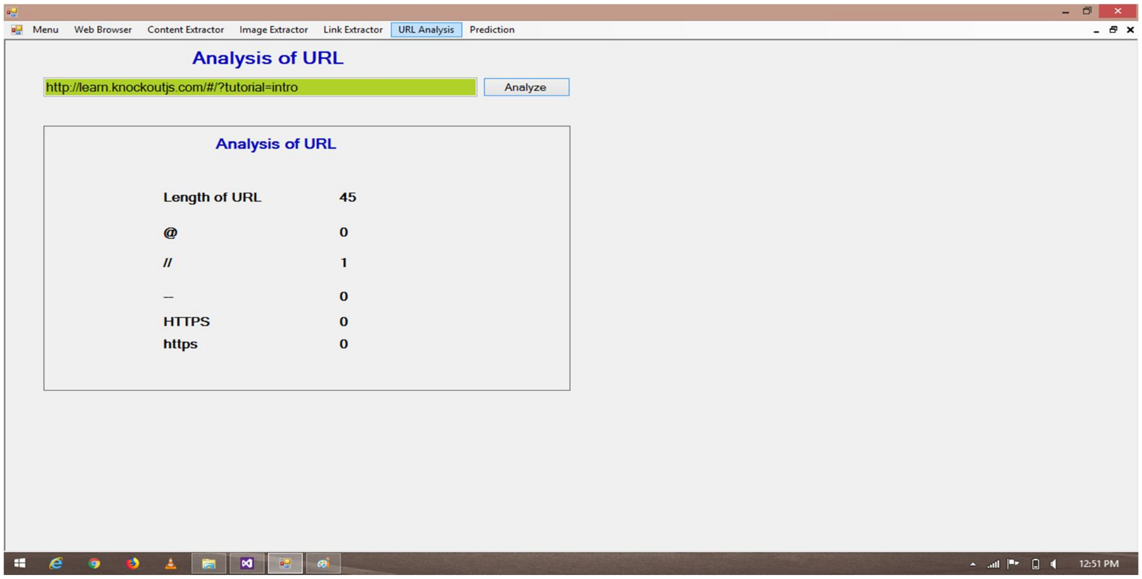
http://learn.knockoutjs.com/#/?tutorial=intro

Total Number of Links : 14

```
http://knockoutjs.com/

http://learn.knockoutjs.com/#
http://knockoutjs.com/






http://knockoutjs.com/downloads/index.html
http://knockoutjs.com/documentation/introduction.html
http://knockoutjs.com/

http://learn.knockoutjs.com/#
http://knockoutjs.com/






http://knockoutjs.com/downloads/index.html
http://knockoutjs.com/documentation/introduction.html
http://knockoutjs.com/

http://learn.knockoutjs.com/#
http://knockoutjs.com/






http://knockoutjs.com/downloads/index.html
http://knockoutjs.com/documentation/introduction.html
```

Fig 4.4 Link Extractor

5) *URL Analysis:* In this diagram URLs are analyzed and shows no of links, symbols, href count.



Analysis of URL

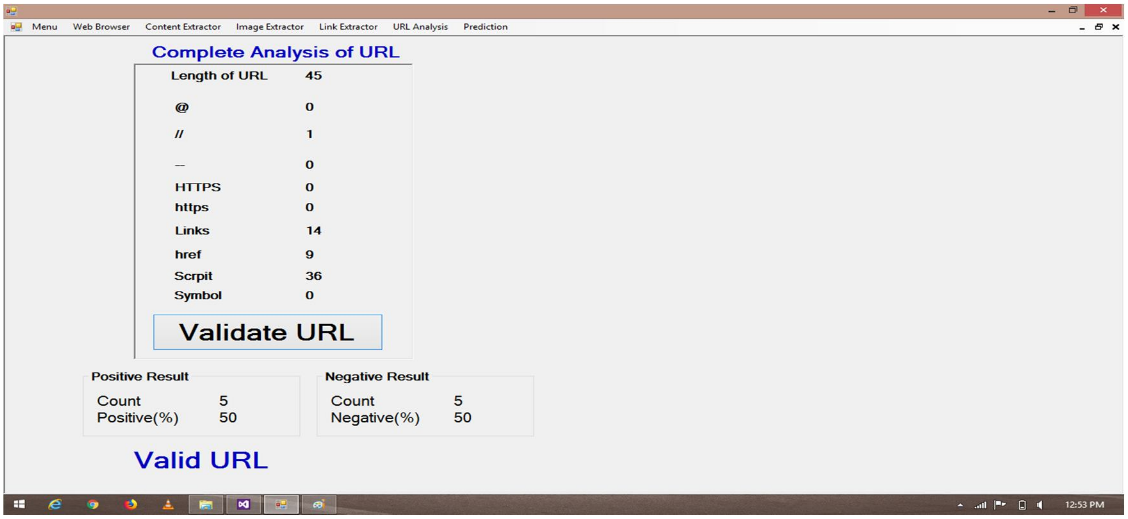
http://learn.knockoutjs.com/#/?tutorial=intro

Analysis of URL

Length of URL	45
@	0
//	1
–	0
HTTPS	0
https	0

Fig 4.5 URL Analysis.

6) *URL Validate:* In this diagram shows the final output of our project. here showing complete analysis of URL and then validate the URL and check the accuracy of URL



Complete Analysis of URL

Length of URL	45
@	0
//	1
–	0
HTTPS	0
https	0
Links	14
href	9
Scrpit	36
Symbol	0

Positive Result

Count	5
Positive(%)	50

Negative Result

Count	5
Negative(%)	50

Valid URL

Fig 4.6 URL Validate.

V. ANALYSIS

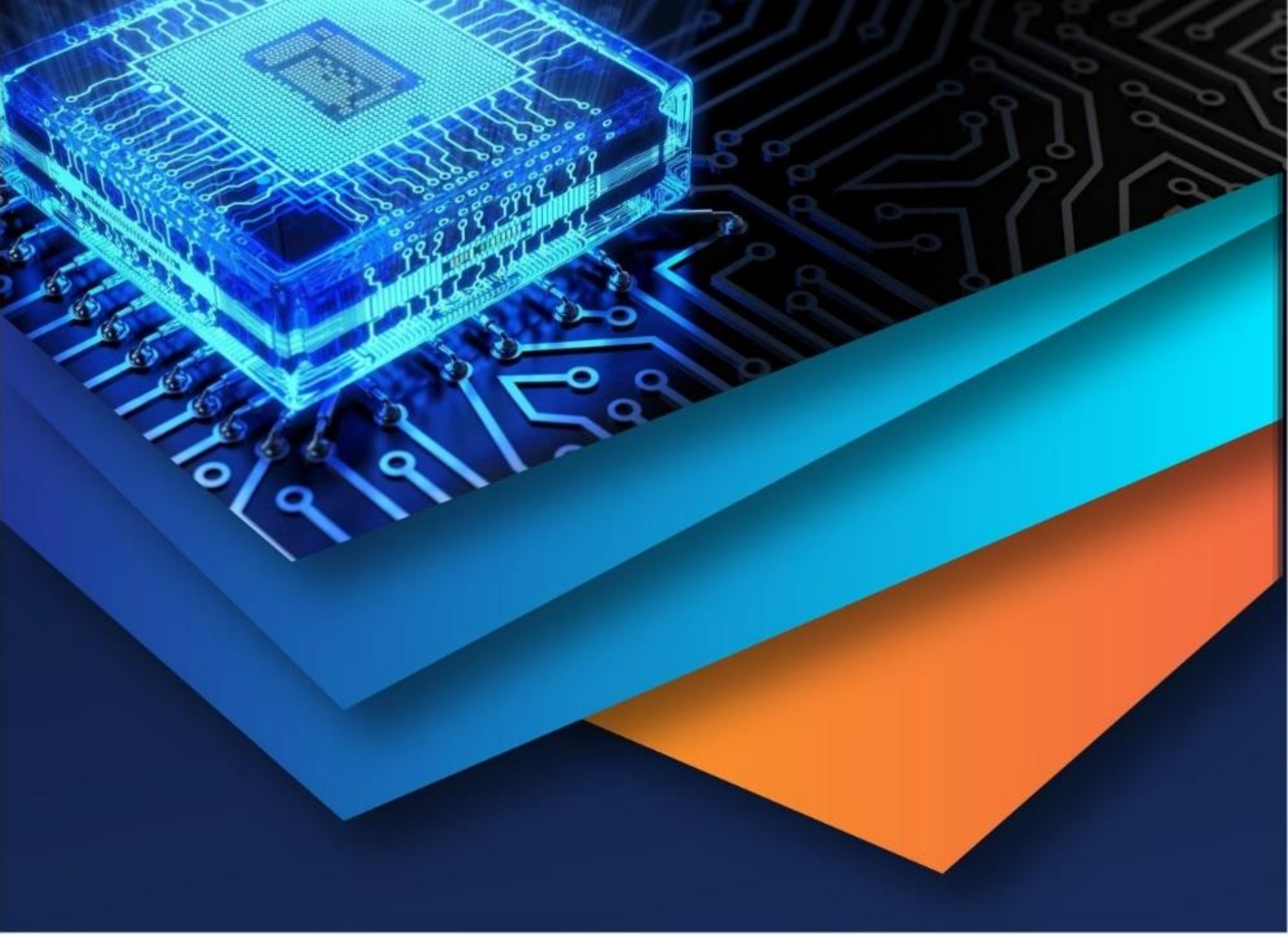
Sr.No	Papers	Techniques	Accuracy
1.	Detecting Phishing Attacks using NLP and Machine Learning	SEAHound Algorithm	Legitimate URL(90% Accuracy)
2.	Fresh Phish :A Framework for auto detection of phishing website	Tensorflow	Legitimate URL(70% Accuracy)
3.	Phishing URL detection machine learning web mining approach	SVM Algorithm	Legitimate URL(65% Accuracy)
4.	Phishstrom: Detection Phishing with Streaming analysis	Search Engine Feature Extraction	Legitimate URL(60% Accuracy)
5.	A novel machine learning and NLP based approach for detecting phishing attacks	NLP & Feature Extraction	Legitimate URL(75% Accuracy)

VI. CONCLUSION

After studying different research papers published nationally and internationally, we have seen many types of techniques. Web browser which successfully browse any kind of websites & work on user behaviour. Our results on phishing URLs demonstrate significantly improved recall which demonstrates that semantic information is a strong indicator of social engineering .Thus the user can check all the contents of web page of that entered URL to decide sharing of personal Information.

REFERENCES

[1] R.kiruthiga, D.Akila, Phishing website detection using machine learning International Journal Of Recent Technology and Engineering, September 2019.
[2] Ram Basnet, Srinivas Mukkamala, Detection of Phishing Attacks: Machine Learning Approach, New Mexico Tech, New Mexico, USA, 2008.
[3] Michael C.Kotson, MIT Lincoln Laboratory Lexington, MA, Characterizing Phishing Threats with Natural Language Processing, 31 Aug 2015.
[4] Dr.G.Ravi Kumar, Dr.S.Gunasekaran, Detection of Phishing Attacks using Machine Learning Techniques, Department of CSE Coimbatore Institute of Engineering,Dec2008.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY