International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023
DOI: 10.55041/ISJEM00384
ISSN: 2583-6129
www.isjem.com
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Detection of Phishing Websites Based on Extreme Machine Learning

Mrs.G..Vijaya lakshmi (Guide)[1], Y.krupajyothi[2], D.Venkata satya sai[3], A.Saikumar[4]

Department of Computer Science & Engineering, Sanketika vidya parisad engineering College Visakhapatnam,AndraPradesh, India.

….. … … … … … … … .. … … … … … .. … … … .. . *******… .. … .. … .. … .. … .. … .. … . … .. … .. … .. … .. .

ABSTRACT

Phishing sites which expects to take the victims confidential data by diverting them to surf a fake website page that resembles a honest to goodness one is another type of criminal acts through the internet and its one of the especially concerns toward numerous areas including e-managing an account and retailing.

Phishing site detection is truly an unpredictable and element issue including numerous components and criteria that are not stable.

On account of the last and in addition ambiguities in arranging sites because of the intelligent procedures programmers are utilizing, some keen proactive strategies can be helpful and powerful tools can be utilized, for example, fuzzy, neural system and data mining methods can be a successful mechanism in distinguishing phishing sites.

Keyword: -

Extreme Learning Machine, Features Classification, Information Security, Phishing.

INTRODUCTION

Phishing is a type of extensive fraud that happens when a malicious website act like a real one keeping in mind that the end goal to obtain touchy data, for example:-passwords, account points of interest, or MasterCard numbers In spite of the fact that there are a few contrary to phishing programming and methods fordistinguishing potential phishing endeavors in messages and identifying phishing substance on sites, phishers think of new and half breed strategies to go around the accessible programming and systems.

Phishing is a trickery system that uses a blend of social designing what's more, innovation to assemble

ISSN: 2583-6129
DOI: 10.55041/ISJEM00384
www.isjem.com

delicate and individual data, for example, passwords and charge card subtle elements by taking on the features of a dependable individual or business in an electronic correspondence.

Existing system

In Phishing E-mail Detection Based on Structural Properties. The proposed approach explains to find phishing through appropriate identification and usage of structural properties of email. The experiment is done by SVM and classification technique to classify phishing e-mails. The technique is used to identify phishing e-mails, which is low in efficiency and scalability.

PROPOSED SYSTEM

The Proposed algorithm is based on automated real-time phishing detection and a machine learning process. The phishing URLs mostly have connections between the part of the URL which means an inter-relation and by using it the features of phishing URLs are extracted.Then the extracted features helps real time detection of phishing websites using machine-learning classification.

SYSTEMANALYSIS

1.Numpy:

Numpyisa general-purpose arra processing package.It provides a high performance multi dimensional array the object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python. It contains various features including

theseimportantones:

- A powerful N-dimensional array object Sophisticated(broadcasting)functions
- Tools for integrating C/C++and For trancode

2.Pandas

Pandasisan open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures.Python was majorly used for data munging and preparation.

ISSN: 2583-6129
DOI: 10.55041/ISJEM00384
www.isjem.com

3.Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in avariety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts,the Python and IPython shells,the Jupyter note book,web application servers, and four graphical user interface tool kits.

4.Scikit–learn

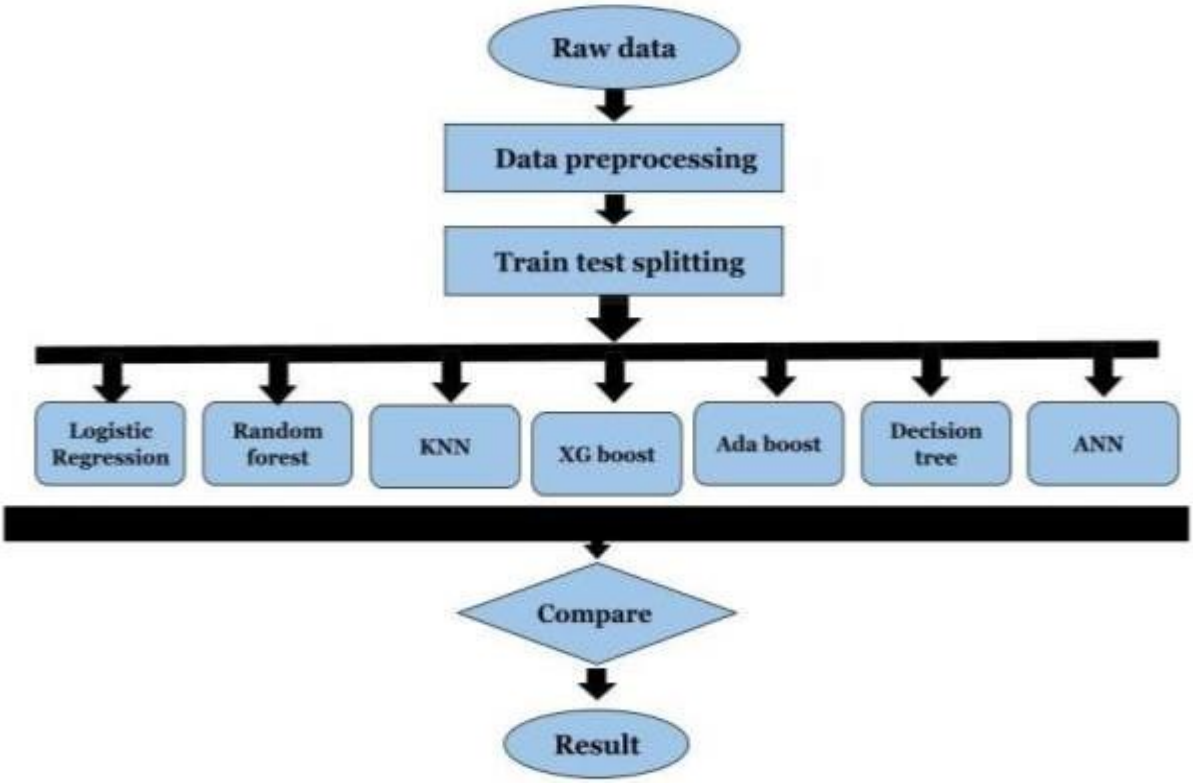
Scikitlearn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python.

METHODOLOGY:

When Object orientation is used in analysis as well as design, the boundary between OOA and OOD is blurred. This is particularly true in methods that combine analysis and design. One reason for this blurring is the similarity of basic constructs (i.e.,objects and classes) that are used in OOA and OOD. Through there is no agreement about what parts of the object- oriented development process belongs to analysis and what parts to design, there is some general agreement about the domains of the two activities.

The main difference between OOA and OOD, due to the different domains of modeling, is in the type of objects that come out of the analysis and design process.

International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00384
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com



FUNCTIONAL REQUIREMENTS

Accuracy of model should be high then only we can get perfect results.Need to be analyze the data and remove the unwanted data, if there is any missing values there need to remove those missing values or else has to put suitable for it.Feature selection is the major part of the data analysis get the perfect feature to build model.

Non Functional Requirements

The major non-functional Requirements of the system are as follows

Usability

The system is designed with completely automated process hence there is no or less user

intervention.

ISSN: 2583-6129
DOI: 10.55041/ISJEM00384
www.isjem.com

Reliability

The system is more reliable because of the qualities that are inherited from the chosen platform java. The code built by using java is more reliable.

Performance

This system is developing in the high level languages and using the advanced front-end and back-end technologies it will give response to the end user on client system with in very less time.

Supportability

The system is designed to be the cross platform supportable. The system is supported on a wide range of hardware and any software platform, which is having JVM, built into the system.

Implementation

The system is implemented in web environment using struts framework. The apache tomcat is used as the web server and windows xp professional is used as the platform.

Interface the user interface is based on Struts provides HTML Tag

International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00384
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

Software Development Life Cycle Model

The Software Development Lifecycle(SDLC) for small to medium database application development efforts.This project uses iterative development lifecycle, where components of the application are developed through a series of tight iteration. The first iteration focus on very basic functionality, with subsequent iterations adding new functionality to the previous work and or correcting errors identified for the components in production.

The six stages of the SDLC are designed to build on one another, taking outputs from the previous stage, adding additional effort, and producing results that leverage the previous effort and are directly traceable to the previous stages

Roles and Responsibilities of PDR AND PER

The iterative lifecycle specifies two critical roles that act together to clearly

communicate project 0issues and concepts between the end-user community and the

development team.

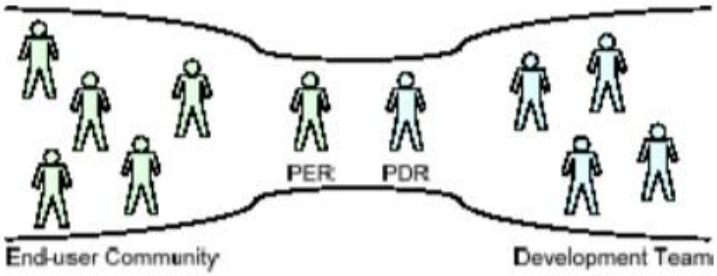
International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00384
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

Primary End-user Representative (PER)

The PER is a person who acts as the primary point of contact and principal

approver for the end-user community. The PER is also responsible for ensuring that appropriate

subject matter experts conduct end-user reviews in a timely manner.



SYSTEM TESTING AND IMPLEMENTATION

Unit Testing:

This testing method considers a module as single unit and checks the unit at interfaces and communicates with other modules rather than getting into details at statement level.

System Testing:

Here all the pre tested individual modules will be assembled to create the larger system and tests are carried out at system level to make sure that all modules are working in synchronous with each other.

Integration Testing

Testing is a major quality control measure employed during software development. Its basic function is to detect errors.

International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00384
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

Top-Down Integration Test

Modules are integrated by moving downwards through the control hierarchy beginning with main program. The subordinate modules are incorporated into structure in either a breadth first manner or depth first manner.

Bottom-Up Integration Test

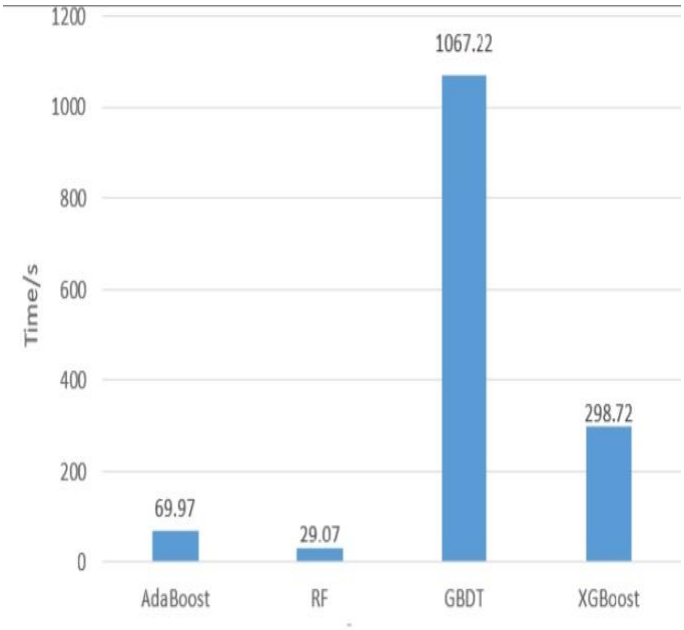
Begins construction and testing with atomic modules. As modules are integrated from the bottom up, processing requirement for modules subordinate to a given level is always available and need for stubs is eliminated.

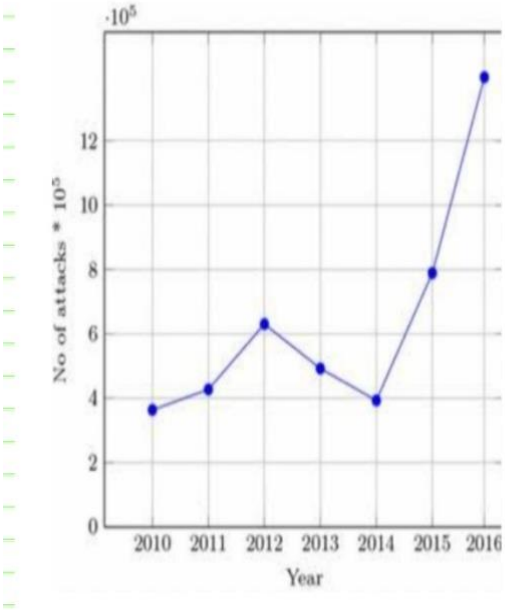
Regression Testing

Regression Testing

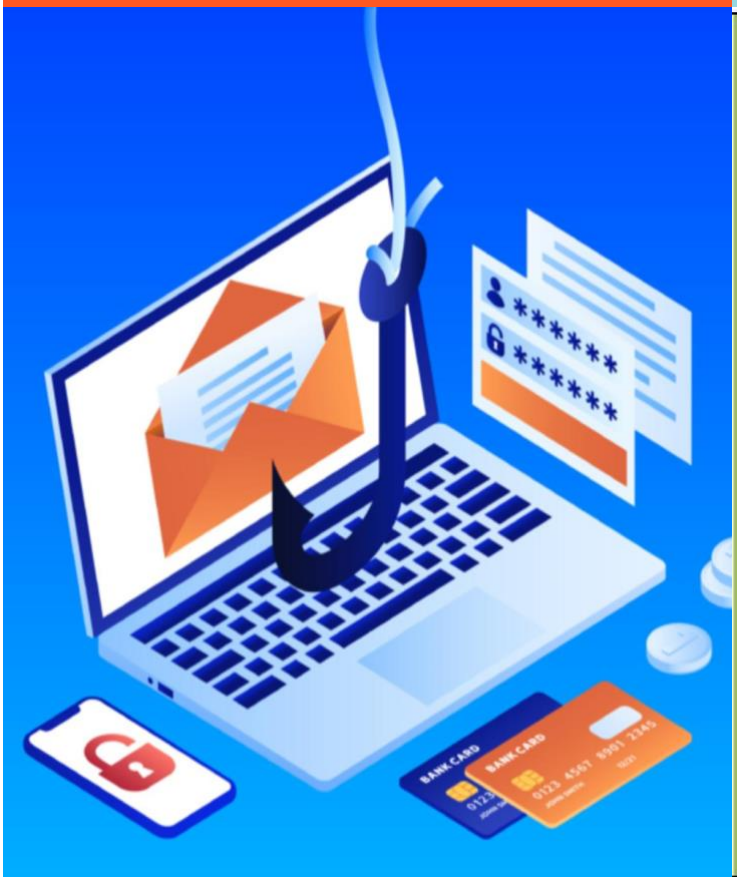
Each time a new module is added as a part of integration as the software changes.

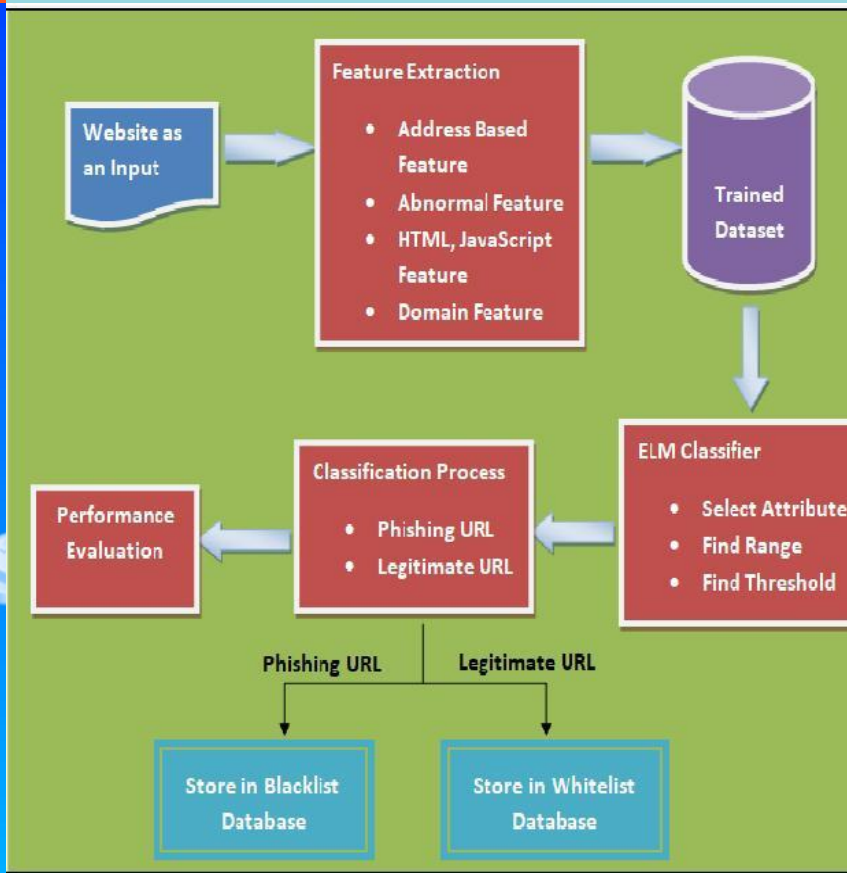
Regression testing is an actually that helps to ensure changes that do not introduce unintended





International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00384
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com





International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00384
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

Source code

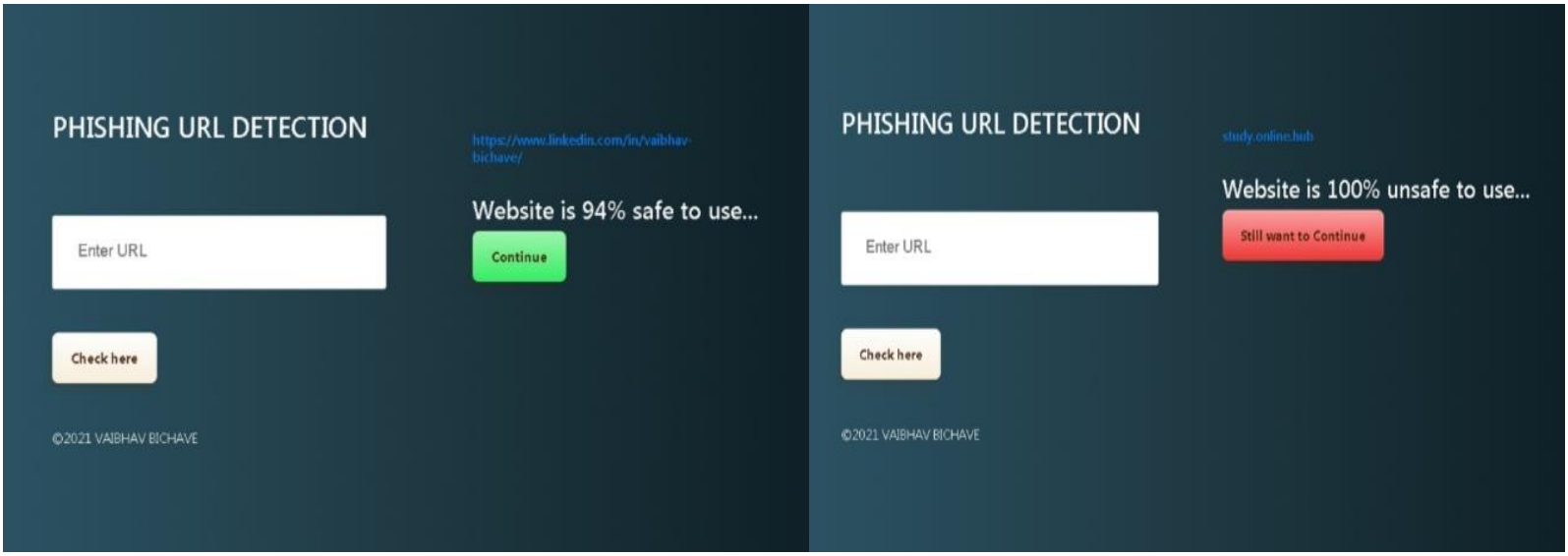
```
1.Phising .py
#!/usr/bin/env python
# coding: utf-8
# In[48]:
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.naive_bayes import GaussianN
from sklearn.metrics import classificatio
from sklearn.metrics import accuracy_scor
# In[47]:
from sklearn.model_selection import train
from sklearn import metrics
from sklearn.svm import SVC
get_ipython().run_line_magic('matplotlib'
sns.set_style('whitegrid')
from sklearn_extensions.extreme_learning_
from sklearn_extensions.extreme_learning_
MLPRandomLayer
```

```
3.Checkurl.py
# -*- coding: utf-8 -*-
import regex
from tldextract import extract
import ssl
import socket
from bs4 import BeautifulSoup
import urllib3.request
import whois
import datetime
def url_having_ip(url):
#using regular function
# symbol =
regex.findall(r'(http((s)?)://)((((\d)+).
# if(len(symbol)!=0):
 # having_ip = 1 #phishing
 # else:
 # having_ip = -1 #legitimate
return 0

def dns(url):
 #ongoing
 return 0
def web_traffic(url):
 #ongoing
 return 0
def page_rank(url):
 #ongoing
 return 0
def google_index(url):
 #ongoing
 return 0
def links_pointing(url):
 #ongoing
 return 0
def statistical(url):
 #ongoing
 return 0
```

International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00384
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

Out put screen shots



CONCLUSION

In this paper, we defined features of phishing attack and we proposed a

classification method consists of feature extraction from websites and classification section. In the feature extraction, we have clearly defined rules of phishing feature extraction and these rules have been used for obtaining

features. In order to classification of these feature, SVM, NB and ELM were used. In the ELM, 6 different activation functions were used and ELM achieved highest accuracy score.

FUTURE SCOPE

The present project is aimed at classification of phishing websites based on the

features. For that we have taken the phishing dataset which collected from uci machine learning repository and we built our model with three different classifiers like SVC, Naïve Bayes, ELM and we got good accuracy scores. There is a scope to enhance it further .if we can have more data our project will be much more effective and we can get very good results. For this we need

API integrations go get the data of different websites.

ISSN: 2583-6129
DOI: 10.55041/ISJEM00384
www.isjem.com

References:

[1] G. Canbek and ù. Sa'Õro'lu, ―A Review on Information, Information Security and Security Processes,‖ Politek. Derg., vol. 9, no. 3, pp. 165– 174, 2006.

[2] L. McCluskey, F. Thabtah, and R. M. Mohammad, ―Intelligent rule- based phishing websites classification,‖ IET Inf. Secur., vol. 8, no. 3, pp. 153–160, 2014. [3] R. M. Mohammad, F. Thabtah, and L. McCluskey, ―Predicting phishing websites based on self-structuring neural network,‖ Neural Comput. Appl., vol. 25, no. 2, pp. 443–458, 2014.

[4]

R. M. Mohammad, F. Thabtah, and L. McCluskey, ―An assessment of features related to phishing websites using an automated technique,‖ Internet Technol. …, pp. 492–497, 2012.

Authors:

Mrs.G.VIjaya Lakshmi is currently working as an Assistant professor in Department of Computer science and Engineering at Sanketika Vidya parishad Engineering college, affiliated to Andhra University. she has more than 2 years of teaching experience .and , Published Papers in Various National & International Journals Her research interest including java,python, Html and.net

Y. Krupajyothi(Team leader) is studying her final year, Bachelor of Technology in sanketika Vidya Parishad Engineering College, affiliated to Andhra University College.As a result of a desire to comprehend the flaws in conventional reporting and to preserve time and high quality. A completely developed project of detection

ISSN: 2583-6129
DOI: 10.55041/ISJEM00384
www.isjem.com

of phising attacks based on machine learning this is used for high security along with code has been submitted for Andhra University as an Academic Project,In completion of the B.tech.

A.sai kumar (Team member) is studying his final year, Bachelor of Technology in sanketika Vidya Parishad Engineering College, affiliated to Andhra University.As academic project,As a result of a desire to comprehend the flaws in conventional reporting and to preserve time and high-quality. A completely developed project of detection of phising attacks based on machine learning this is used for high security along with code has been submitted for Andhra University as an Academic Project. In completion of the B.tech.

D.Venkata sathya sai (Team member) is studying his final year, Bachelor of Technology in sanketika Vidya Parishad Engineering College, affiliated to Andhra University.As academic project,As a result of a desire to comprehend the flaws in conventional reporting and to preserve time and high-quality. A completely developed project of detection of phising attacks based on machine learning this is used for high security along with code has been submitted for Andhra University as an Academic Project. In completion of the B.tech.