RESEARCH ARTICLE OPEN ACCESS

Improving Intelligent Phishing Website Detection using Particle Swarm Optimization

Hemanjali.N [1], Krishna Satya Varma .M [2]
Department of IT, S.R.K.R Engineering College, Bhimavaram - India

ABSTRACT
Web phishing assaults have been evolving over the last few years, causing users to lose trust in e-commerce and online businesses. To detect phishing websites, several techniques and systems based on a blacklist of phishing websites are used. Unfortunately, the rapid advancement of technology has resulted in the development of more complex strategies for attracting consumers to websites. As a result, current blacklist-based techniques are unable to detect the most recent and newly launched phishing websites, such as zero-day phishing websites. Machine learning algorithms have been used in several recent research papers to identify phishing websites and to use them as an early warning system to identify such risks. In most of these approaches, however, the significant website features have been chosen based on human experience or frequency analysis of website features. Phishing website detection using particle swarm optimization-based feature weighting is proposed in my study to improve phishing website detection. Using the PSO technique and five machine learning algorithms, we classify whether the provided URL is phishing or not. As a result, we use the five machine learning algorithms to detect phishing websites. To obtain improved accuracy for detecting phishing websites, the proposed approach involves using particle swarm optimization (PSO) to appropriately weight numerous website attributes. The experimental findings suggested that the proposed PSO-based feature weighting improved classification accuracy significantly.

Keywords:-phishing website detection, Particle Swarm Optimization, Machine learning algorithms, feature weighting

I. INTRODUCTION

In today's cyber world, the majority of individuals communicate with one another using a computer or a digital gadget connected to the internet. Because of the convenience, comfort, and assistance that e-banking, online shopping, and other online services provide, the number of people who use them is growing. An attacker sees this instance as an opportunity to make money or gain notoriety, thus they steal crucial information required to access internet service websites. Phishing is one of the methods for obtaining sensitive data from people. It's sent through a spoof page on a legitimate website, tricking online consumers into giving personal information. Phishing is a phrase that refers to the act of 'fishing' for sensitive information from victims. The attacker sends a bait in the form of a spoof website and waits for sensitive data to be returned. Phone phreaking, a common technique for illegally exploring telephone lines, has inspired the replacement of the 'f' phoneme with the 'ph' phoneme. The attacker succeeds when he convinces a victim to trust the phoney page and obtains the victim's credentials for the imitated authentic site. Threat Cyber Group, Internet Identity (IID), Mark Monitor, Panda Security, and Forcepoint are members of the Anti-Phishing Working Group (APWG), a non-profit organisation that investigates phishing assaults reported by its members. It investigates the attacks and provides reports on a regular basis. It also provides statistics on malicious domains and phishing assaults that are taking place around the world. Phishing attracts online users for a variety of reasons: 1.Lack of knowledge about computer systems. 2.Lack of knowledge about security and security indicators.
3.Lack of awareness about the caution and progress by reducing the strength of the existing tools. 4.lack of awareness about the visual text that fraud by making similar url and website content.

Phishing attacks can take many forms, including email, websites, and malware. In order to carry out email phishing, attackers create bogus emails that appear to come from a reputable company[1]. They send fake emails to a large number of online users, thinking that thousands of genuine people will fall for it. Phishing is the process of someone attempting to obtain your personal information, such as your passwords, credit card number, checking account information, or other information protected under the Data Protection Act. Such attempts, which are commonly referred to as Phishing assaults, are frequently crude and visible. Please keep in mind, though, that they are becoming more sophisticated. In website phishing, the attacker creates a website that looks to be a clone of a legitimate website and lures users to it via adverts on other websites or social media sites like as Facebook and Twitter. Several attackers can control phishing websites with security indicators such as a green padlock, HTTPS connection, and so forth. Malware phishing occurs when an attacker injects harmful software, such as a computer virus, into a compromised legitimate website without the victim's awareness. According to the APWG study, there were 20 million new malware samples detected. The vast majority of recent malware is multifunctional, which means it can steal data, make the victim's system a botnet node, or download and install other malicious software without the client's knowledge.Spear-phishing is a type of phishing attack that targets a specific group of people or a community within a company or enterprise. To detect and halt phishing assaults, numerous anti-phishing approaches have been proposed in the literature[2]-[5].

II. LITERATURE REVIEW

Here we review a number of the prevailing work that has been implemented by many authors regarding phishing sites using machine learning techniques.In [6] authors implemented ‘‘Intelligent phishing detection and protection scheme for online transactions’’. to resolve this problem , the present study introduces advanced inputs such as Legitimate site rules, User-Behaviour profile, Phish tank, User-specific sites,Pop-Ups from emails which weren’t considered previously during a single protection platform. The idea is to employ a Neuro-Fuzzy Scheme with five inputs to detect phishing sites in real time with great accuracy. The suggested model is trained and tested using 2-fold cross-validation in this study. a complete of 288 features with 5inputs were used . In [7] Qabajeh and F. Thabtah proposed ‘‘An experimental study for assessing email classification attributes using feature selection methods’’. He suggested using the IG, Chi-square, and Correlation Features Set (CFS) to scale back the data dimensionality and choose the minimal set of important features. In [8] author proposed Detection of phishing attacks. In this type of cyber attack, the attacker sends malicious links or attachments through phishing e-mails that can perform various functions, including capturing the login credentials or account information of the victim. In this study, a software called “Anti Phishing Simulator” was developed, giving information about the detection problem of phishing and how to detect phishing emails. With this software, phishing and spam mails are detected by examining mail contents. Classification of spam words added to the database by Bayesian algorithm is provided.In [9] F. Thabtah and N. Abdelhamid proposed an intelligent approach as "Deriving correlated sets of Website features for phishing detection: A computational intelligence approach". He employed Information Gain (IG), Chi-square, and Correlation Feature Set to seek out the foremost significant website's features to reinforce the detection accuracy of phishing websites for a few rule-based classification machine learning algorithms: C4.5, RIPPER, and PART. In[10] Phishing attacks are common online, which have resulted in financial losses through using either malware or social engineering. Thus, phishing email detection with high accuracy has been a problem of great interest. Machine learning-based detection methods, particularly Support Vector Machine (SVM), are proved to be effective. However, the parameters of the kernel method, whose default is that class numbers reciprocals generally , affect the classification accuracy of SVM. to enhance the classification accuracy, this paper proposes a model, called Cuckoo Search SVM(CS-SVM). The CS-SVM extracts 23 features, which are accustomed construct the hybrid classifier. within the hybrid classifier, Cuckoo Search (CS) is integrated with SVM to optimize the parameter selection of the Radial Basis Function(RBF). An evaluation is carried out on a dataset consisting of 1,384 phishing emails and 20,071 non-phishing emails.

III. EXISTING SYSTEM

A.Heuristic based technique:

To detect phishing attacks, these strategies leverage features derived from the phishing website. Since certain phishing sites lack common characteristics, this approach has a low detection rate. As this method does not rely on a list-based comparison, it produces fewer false positives and negatives. This approach detects zero-day phishing assaults that were missed by list-based strategies. The disadvantages of this include -that it is less accurate than list-based strategies because there is no guarantee that these qualities will be present on all phishing websites and also that once an attacker understands the algorithm or features used in detecting phishing sites, he can circumvent the heuristic features and achieve his goal of stealing sensitive information.

B.Visual Similarity-based Approach

The main objective of the phisher is to deceive the user by designing a particular image of a legitimate site such the user doesn't get any suspicion of the phishing site. Compare suspicious website images with legitimate image databases to urge the similarity ratio. Disadvantages are-

1.Image comparison of the suspicious website with entire legitimate database store takes longer complexity.
2.Huge space to store valid image database.

C. Machine learning-based techniques

Nowadays, most researchers are concentrating on the utilization of machine learning algorithms (ML). But these machine learning algorithms may cause poor detection accuracy and time consuming because it iterates on an outsized set of the dataset. a number of the machine learning algorithms are sequential minimum optimization (SMO), J48 tree, Random Forest (RF), logistic regression (LR), multilayer perceptron (MLP), Bayesian network(BN), support vector machine (SVM) and AdaBoostM1 .Disadvantage of this is-

a huge amount of data in many fields resulted in computational challenge,which makes machine learning algorithms difficult to applied in real-world problems

IV. PROPOSED SYSTEM

• Use machine learning algorithms to combine new heuristic features to reduce false positives when detecting new phishing sites.

• Machine learning algorithms augmented by the proposed PSO-based feature weighting provide higher detection accuracy and outperform stand-alone machine learning models when other feature selection methods are used. PSO's most significant advantage over other optimization algorithms is its ability to achieve rapid convergence in a wide range of complex optimization situations. Furthermore, PSO has a number of appealing features, including reduced mathematical equations and fewer parameters in implementation.

• The PSO-based website feature weighting is used to distinguish between various website features based on how relevant they are in distinguishing phishing from legitimate websites.

A.Advantages

1. Accurately detects the phishing website
2. Feature selection decreases learning time and improves reliability.
3. Machine learning has a high level of accuracy.

V. SYSTEM IMPLEMENTATION

A.Datacollection

The phishing websites dataset available in UCI Machine Learning Repository to guage the performance of the proposed PSO-based feature weighting approach suggested improving the phishing website detection [12]. This dataset contains several legitimate and phishing websites. of these data are sent to the feature extraction module to extract the important features from the data. This dataset contains 4898 phishing websites and 6157 legitimate websites out of 11055 websites. The below table represents the characteristics of the phishing websites datasets utilized in the evaluation.

TABLE I
PHISHING WEBSITES DATASET

Characteristic	Description
Websites features	12 features of address bar based category, 6 features of abnormality-based category, 5 features of HTML and JavaScript-based category, and 7 features of domain-based category
Features	30
Classes	Legitimate or phishing website
Websites	11055
Phishing Websites	4898
Percentage of phishing websites	44%

Legitimate websites	6157
Percentage of legitimate websites.	56%

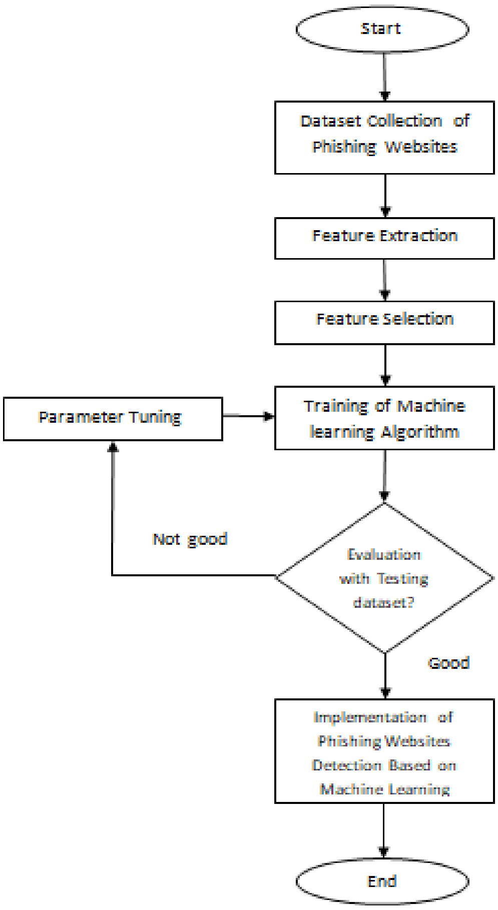

Start
Dataset Collection of Phishing Websites
Feature Extraction
Feature Selection
Parameter Tuning
Training of Machine learning Algorithm
Not good
Evaluation with Testing dataset?
Good
Implementation of Phishing Websites Detection Based on Machine Learning
End

FIG 1 DATA FLOWCHART

B. Modules

Here we use a building block of modules to detect phishing websites. Module start with the data collection of the websites .

1)Initialize Dataset: To implement a model the primary thing we require is to initialize the dataset because the model works on a dataset. For intelligent detection of PSO- based phishing websites, we want a raw dataset of 11055 URL websites that contains the amount of legitimate and phishing websites. these raw datasets are further extracted and processed to train the model.

2)Feature Extraction: Feature extraction could be part of the dimensionality reduction process, which involves splitting and reducing a large set of data into smaller groupings. So when you're ready to process, it'll be a lot easier. An intelligent phishing website detection strategy is based on identifying common information from websites to efficiently train machine learning models to recognise phishing websites. Some elements of websites are more important and essential than others when it comes to distinguishing phishing from authentic websites. As a result, obtaining these discriminative website elements is critical for improving phishing website detection accuracy. In order to identify the most important aspects in phishing website detection, a number of prominent website features are studied and assessed in the literature. The authors of [11] divided important phishing website features into four categories: HTML and JavaScript-based features, address bar-based features, domain-based features, and abnormality-based features. Initially, 12 features from the address bar, 6 features from abnormalities, 5 features from HTML and JavaScript, and seven features from a domain-based category are taken from 11055 websites.

3) Data Preprocessing: To improve phishing website prediction, the PSO-based feature weighting strategy is commonly recommended in data processing.Feature weighting is a common approach for determining the best weight for each feature, which symbolises the feature's value in the classification decision. To improve the performance of machine learning models, it seeks to assign lower weights to less influential features and higher weights to more significant features.For feature weighting, we employ the particle swarm optimization technique because it is easier to implement and contains less mathematical equations. To improve the effectiveness of phishing website identification, the simplest weights of website features are heuristically constructed using PSO.

4)Training: A group of phishing and legitimate websites are used to train and construct an intelligent detection model during the training phase. After that, the feature extraction training phase is carried out, which uses the

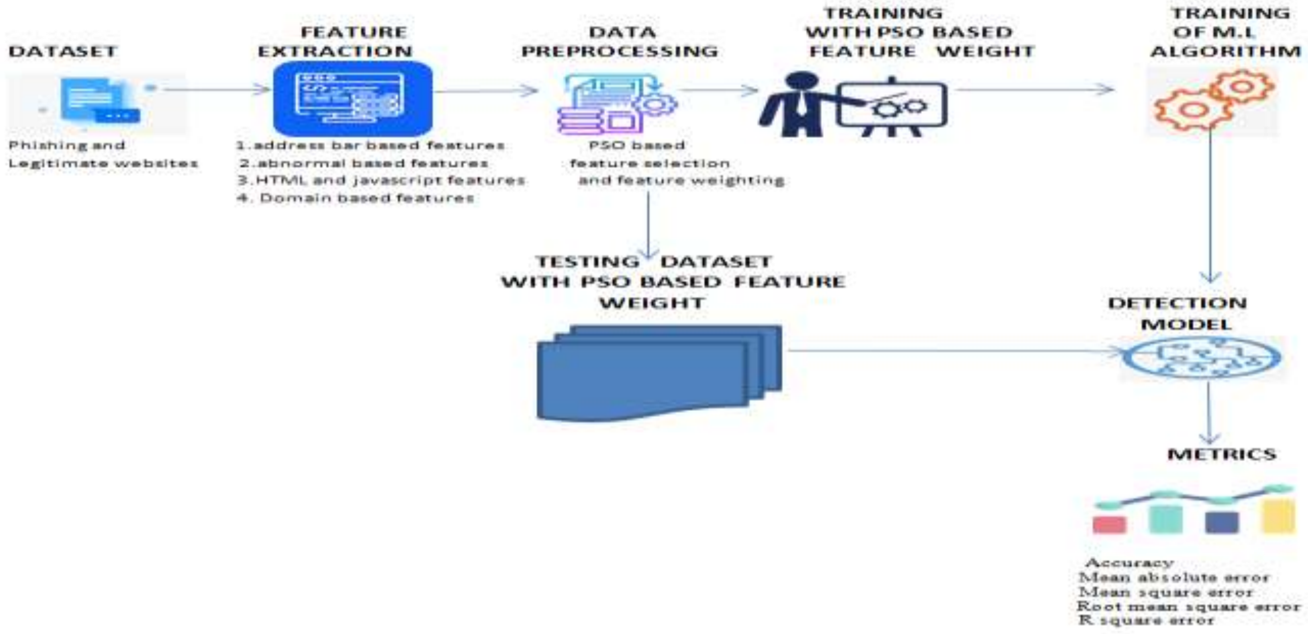


FIG-2 SYSTEM ARCHITECTURE

feature extracted dataset as input. The suggested PSO-based feature weighting is used to weight these retrieved website properties . Following feature weighting based on PSO, machine learning algorithms are trained to detect phishing websites using the features weighted websites. The training dataset of features weighted by PSO is used to train KNN, random forest, and SVM machine learning algorithms. After that, the trained models are constructed and saved, ready to be used in the detection phase to detect new phishing websites.

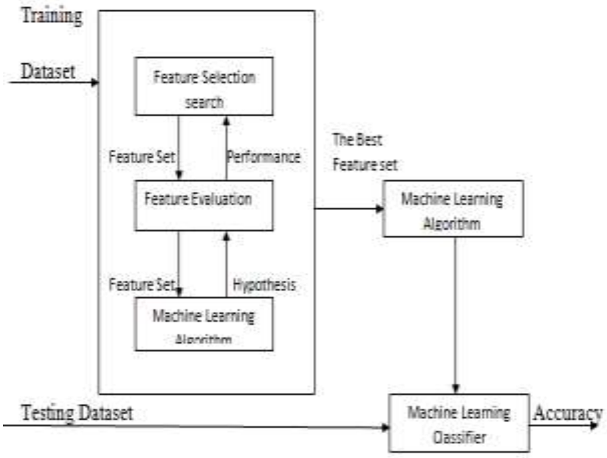


FIG-3 FEATURE SELECTION APPROACH USED FOR PREDICTING THE PHISHING WEBSITE.

5)Testing: New websites are collected and utilised as a testing dataset in the detection phase to evaluate the efficiency of the phishing website detection models generated in the training phase. The ideal weights acquired by PSO during the training phase are then used to weight aspects of the testing dataset in order to improve phishing website detection accuracy. As a result, the PSO-weighted features are utilised as inputs to the phishing website detection models developed during the training phase, which classify whether the website is phishing or not.

VI. METHODOLOGY

Here we use the particle swarm optimization method for data pre-processing to extend the performance of machine learning. PSO-based website feature weighting is employed to differentiate between the various features of internet sites , supported how important they contribute towards recognizing phishing from legitimate websites.

I. PARTICLE SWARM OPTIMIZATION

Every candidate solution in PSO is represented by a particle that flies with a specific velocity during a swarm or population of alternative solutions. All particles in the swarm are initially formed by assigning random positions and velocities to them. Then, each particle in the swarm dynamically modifies its position and velocity in accordance with its own and its friends' flying experiences. Each particle keeps a record of its previous best position (Pbest) in each PSO iteration and has access to the global best position (gbest). To achieve the best solution inside the swarm, each particle modifies its position and velocity based on pbest and gbest using Equations (1) and (2).

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \quad (1)$$

$$v_{id}^{t+1} = w * v_{id}^{t} + c_1 * r_1 * (p_{id} - x_{id}^{t}) \quad (2)$$

Xid ^t and vid^t denote the location and velocity of particle I at iteration t, respectively, with d = 1, 2, 3... D (D is that the dimensionality of search space). pid and pgd stand for pbest and gbest, respectively. Learning rates are denoted by the positive constants c1 and c2, which are typically set to 2.0.They define the weighting of the stochastic acceleration terms that pull each particle to its pbest and gbest places. w represents the inertia weight whereas r1 and r2 are randomly set to real numbers within the interval [0, 1].

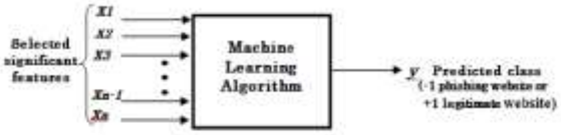


FIG-4 INPUT AND OUTPUT OF THE MACHINE LEARNING CLASSIFIERS USED FOR PREDICTING THE PHISHING WEBSITE

Here we compare the result using machine learning algorithms to train the model using the training dataset with the features weighted by PSO.

II Random Forest Model

Random Forest is a well-known machine learning method that is used in supervised learning. As the name suggests, "Random Forest" is a classifier that employs numerous decision trees on distinct subsets of a dataset and utilises the average to increase the dataset's predictive accuracy.

The Working process can be described in the below steps :

Step-1: Pick K data points at random from the training set.

Step-2: Create decision trees based on the data points you've chosen (Subsets).

Step-3: Select the number N for the number of decision trees you wish to make.

Step-4: Repetition of Steps 1 and 2.

Step-5: Find the forecasts of each decision tree for new data points, and allocate the new data points to the category that receives the most votes.

III SUPPORT VECTOR MACHINE

The support vector machine algorithm's goal is to find a hyperplane in N-dimensional space (N — the number of characteristics) that categorises the data points clearly. Hyper-planes are decision-making boundaries that aid in data categorization.Following are the steps for SVM

Step1:- We must first determine the correct hyperplane.

Step2:- The second phase is to maximise the distances between neighbouring data points after the primay step.

Step3:- If the data is non-linear, add a feature z=x^2+y^2.

Step4:- To classify the category, use the SVM classifier.

The classification is binary.

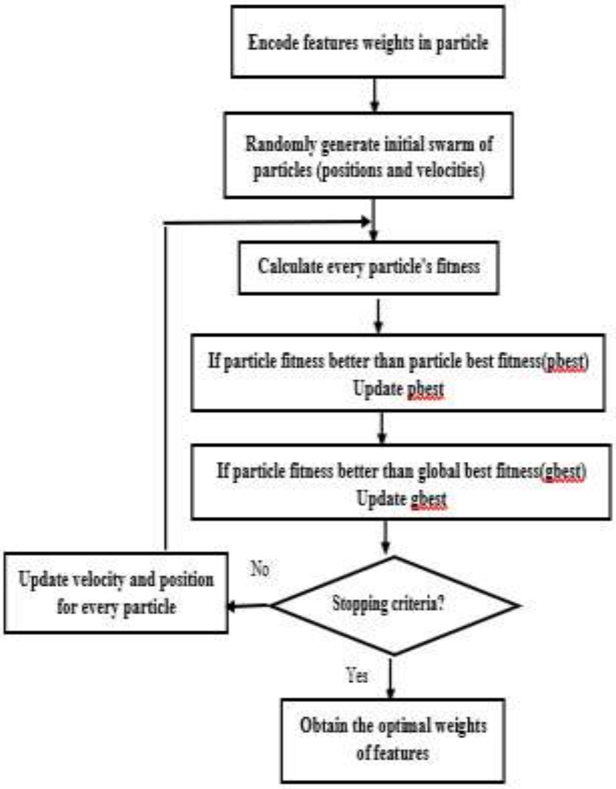

Encode features weights in particle
Randomly generate initial swarm of particles (positions and velocities)
Calculate every particle's fitness
If particle fitness better than particle best fitness(pbest) Update pbest
If particle fitness better than global best fitness(gbest) Update gbest
Stopping criteria?
No
Update velocity and position for every particle
Yes
Obtain the optimal weights of features

FIG-5 PSO ALGORITHM FLOW CHART

IV K- NEAREST NEIGHBOUR

Step-1: Determine the number K of neighbours.

Step-2:Determine the Euclidean distance between K neighbours.
Step-3: Using the obtained Euclidean distance, find the K closest neighbours.
Step-4: Count the number of data points in each category among these k neighbours.
Step-5: Assign the new data points to the category with the greatest quantity of neighbours.
Step-6: We've completed our model.

Euclidean distance(d) = $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

V LOGISTIC REGRESSION

In simple, linear regression, predict scores on one variable from the scores on a second variable. The variable that's predicted is named the criterion variable and is mentioned as Y. The variable base for predictions is named the predictor variable and is mentioned as X. When there's just one predictor variable, the prediction method is named simple regression. In simple linear regression, the subject of this section, the predictions of Y when plotted as a function of X form a straight line.

VI DECISION TREE

In classification and regression, Decision Trees (DTs) are a non-parametric supervised learning method. The goal is to create a model that uses simple decision rules inferred from data attributes to forecast the value of a target variable. The following is how the decision tree works:
Step1: At the root of the tree, place the dataset's basic attribute.
Step 2:Split the training set into subsets. Subsets should be made in such a way that every subset contains data with the equivalent.
Step 3:Repeat step 1 and step 2 on each subset until you discover leaf nodes altogether the branches of the tree. When using decision trees to forecast a record's class label, we start at the bottom of the tree. The values of the base attribute are then compared to the value of the record's attribute. Follow the branch related to that value and jump to the next node based on the comparison.

VII. RESULT AND ANALYSIS

Implemented accuracy of 5 machine learning algorithms on the given dataset using particle swarm optimization for phishing website detection shows that the accuracy of Random forest is high compared to other models.

Accuracy is that the metric that counts the proportion of correct predictions.

ACCURACY = $\frac{TP+TN}{TP+TN+FP+FN}$

TABLE II
COMPARISION OF ACCURACY ACHIEVED BY MACHINE LEARNING ALGORITHMS BEFORE AND AFTER APPLYING PSO

Algorithm	Accuracy without PSO	Accuracy with PSO-based feature selection
KNN	93.71	95.21
SVM	89.38	94.1
Logistic Regression	90.21	93.51
Decision Tree	88.50	91.72
Random Forest	94.81	97.31

The above table represents the accuracy of machine learning algorithms KNN, SVM, Logistics regression, Decision tree, and Random forest for phishing website detection without PSO- based feature selection and with PSO -based feature selection.

I Error evaluation

In the proposed Intelligent Phishing Website Detection using Particle Swam Optimization, we use 4 metrics for the error evaluation of the results.

- Mean absolute error
- Mean square error
- Root mean square error
- R square error

A.Mean absolute error

It calculates the average magnitude of mistakes in a group of forecasts without taking into account the direction of the errors. It's the average of the absolute differences between forecast and actual observation over the test sample, where all individual deviations are given equal weight.

$$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n}\left|y_j - \hat{y}_j\right|$$

B.Root mean square error

It’s a quadratic scoring rule that additionally calculates the error's average magnitude. It's the square root of the average of squared discrepancies between predicted and observed values

$$.\text{RMSE} = \sqrt{\frac{1}{n}\Sigma_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2}$$

C. Mean square error

It is the typical of the squared error that's used because the loss function for method of least squares regression. it's the sum, overall the info points, of the square of the difference between the anticipated and actual target variables, divided by the amount of knowledge points.

$$\text{MSE} = \frac{1}{n}\Sigma_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

D. R square error

To determine how well the regression curve fits the info , we discover a worth called R-Squared (r2) to seek out r2, simply square the correlation. The closer r2 is +1, the higher the road fits the info r2 will always be a positive number.

$$R^2 = 1 - \frac{\textit{sum of squared distances between the actual and predicted } Y \textit{ values}}{\textit{sum of squared distances between the actual } Y \textit{ values and their mean}}$$

TABLE III
AVERAGE VALUE OF ERROR EVALUATION
METRICSACHIEVED BY MACHINE LEARNING ALGORITHMS
USING PSO-BASED FEATURE SELECTION

Algorithm	Mean square error	Mean absolute error	R-squared error	Root mean square error
Support vector machine	0.337192	0.168596	0.657313	0.580683
Random forest	0.125904	0.062952	0.872919	0.35483
Logistic regression	0.322721	0.16136	0.67086	0.568085
K-Nearest neighbor	0.277858	0.138929	0.714368	0.527123
Decision tree	0.345876	0.172938	0.650792	0.588112

II EXECUTION

A .Main.py

```
import Tkinter as tk
from Tkinter import Message, Text
from PIL import Image, ImageTk
import pandas as pd

import tkinter.ttk as ttk
import Tkinter. font as font
import tkinter.messagebox as tm
import matplotlib.pyplot as plt
import csv
import numpy as np
from PIL import Image, ImageTk
from Tkinter import filedialog
import tkinter.messagebox as tm
import RandomForest as RF
import LogisticRegression as LR
import SVM as SV
import DecisionTree as DT
import KNN as knn
from sklearn.externals import joblib

bgcolor="#DAF7A6"
bgcolor1="#B7C526
fgcolor="black"
def Home():
        global window
        def clear():
          print("Clear1")
          txt.delete(0, 'end')
          txt1.delete(0, 'end')
```

```
            window = tk.Tk()
             window.title("Detecting Phising Website")
            window.geometry('1280x720')
             window.configure(background=bgcolor)
             #window.attributes('-fullscreen', True)
           window.grid_rowconfigure(0, weight=1)
             window.grid_columnconfigure(0, weight=1)
             message1 = tk.Label(window, text="Detecting
Phising Website" ,bg=bgcolor  ,fg=fgcolor  ,width=50
,height=3,font=('times', 30, 'italic bold underline'))
             message1.place(x=100, y=20)
            lbl = tk.Label(window, text="Select
Dataset",width=20  ,height=2  ,fg=fgcolor  ,bg=bgcolor
,font=('times', 15, ' bold ') )
             lbl.place(x=100, y=200)
```

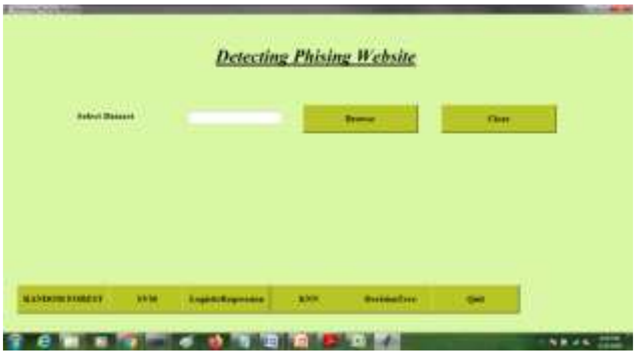
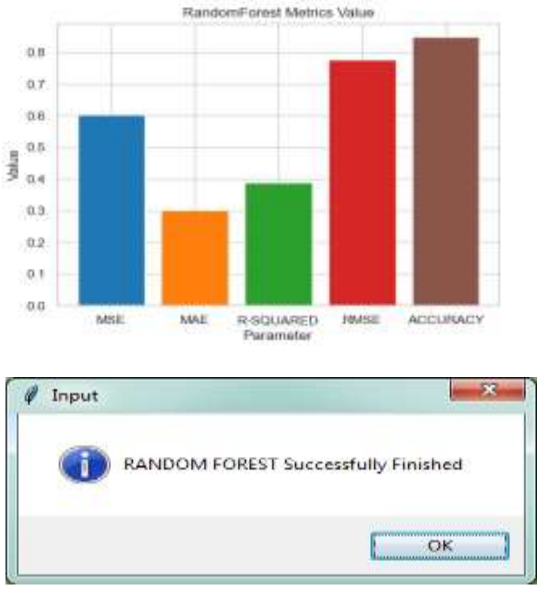
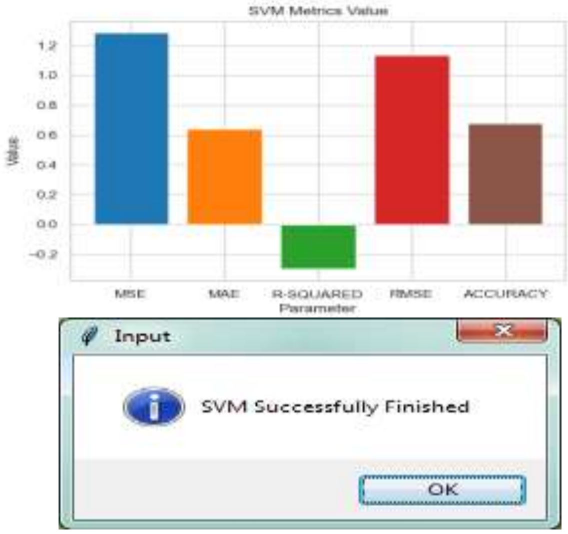


FIG-6 APPLICATION HOME PAGE



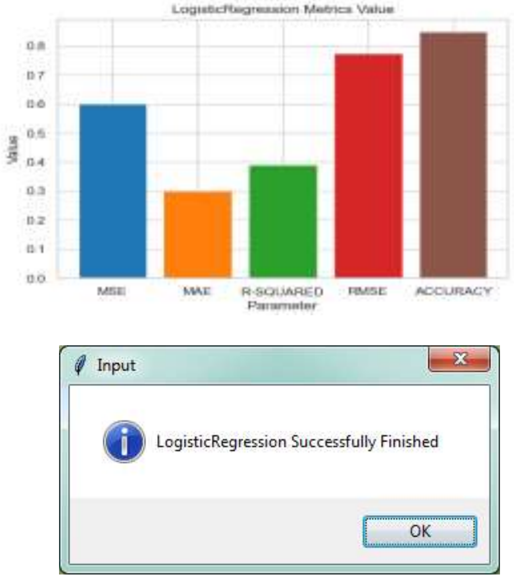
RANDOM FOREST Successfully Finished

FIG-7 EVALUATION METRICS ARRIVE THROUGH RANDOM FOREST



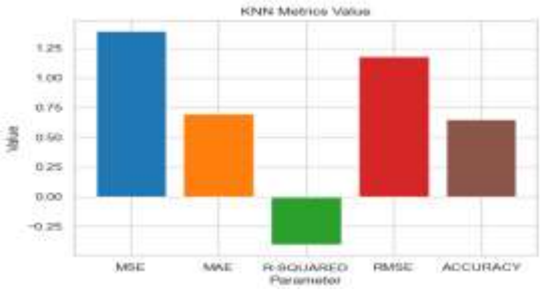
SVM Successfully Finished

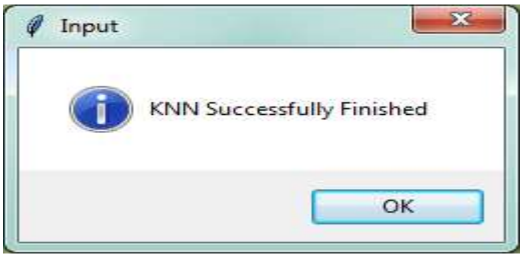
FIG-8 EVALUATION METRICS ARRIVE THROUGH SVM



LogisticRegression Successfully Finished

FIG-9 EVALUATION METRICS ARRIVE THROUGH LOGISTIC REGRESSION





KNN Successfully Finished

FIG-10 EVALUATION METRIC ARRIVE THROUGH KNN

VIII. CONCLUSION

Phishing is a type of cybercrime that uses social engineering and specialised deception to get personal information. Phishing, on the other hand, is a broad category of deception. Experiments with recent credible phishing data sets using various categorization algorithms with various learning methodologies are carried out. the base of the experiments is accuracy measure. This research work aims to predict whether a given URL is a phishing website or not then compare the accuracy results before PSO and after PSO. The web site features were weighted with the best weights by applying PSO to reinforce the detection of phishing websites in the proposed PSO-based feature weighting. We implemented five different machine learning algorithms and determine the most effective one.

REFERENCES

[1] Sadeh N, Tomasic A, Fette I. "Learning to detect phishing emails". *Proceedings of the 16th international conference on World Wide Web*. 2007: p. 649-656.

[2]W. Ali and A. A. Ahmed, ''Hybrid intelligent phishing Website prediction using deep neural networks with genetic algorithm-based feature selection and weighting,'' *IET Inf. Secure*., vol. 13, no. 6, pp. 659–669, Nov. 2019.

[3] Andr Bergholz, Gerhard Paa, Frank Reichartz, Siehyun Strobel, and Schlo Birlinghoven Improved phishing detection using model-based features *In Fifth Conference on Email and Anti-Spam*, CEAS, 2008

[4]R. M. Mohammad, F. Thabtah, and L. McCluskey, ''Tutorial and critical analysis of phishing Websites methods,'' *Comput. Sci. Rev*., vol. 17, pp. 1–24, Aug. 2015.

[5]S. Nawafleh, W. Hadi (2012). Multi-class associative classification to predicting phishing websites. *International Journal of Academic Research Part A;* 2012;4(6), 302-306J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[6] P. A. Barraclough, M. A. Hossain, M. A. Tahir, G. Sexton, and N. Aslam, ''Intelligent phishing detection ad protection scheme for online transactions,'' *Expert Syst. Appl.*, vol. 40, no. 11, pp. 4697–4706, 2013.

[7] I. Qabajeh and F. Thabtah, ''An experimental study for assessing email classification attribute using feature selection methods," *in Proc. 3rd Int. Conf. Adv. Comput. Sci. Appl. Technol.* (ACSAT), Dec. 2014, pp. 125–132.

[8] M. Baykara and Z. Z. Gürel, "Detection of phishing attacks," *6th International Symposium on Digital Forensic and Security* (ISDFS), Antalya,2018, pp. 1-5.

[9]F. Thabtah and N. Abdelhamid, *''Deriving correlated sets of Website features for phishing detection: A computational intelligence approach,''* J. Inf. Knowl. Manag., vol. 15, no. 4, pp. 1–17, 2016.

[10]W. Niu, X. Zhang, G. Yang, Z. Ma, and Z. Zhuo, "Phishing Emails Detection Using CS-SVM," *IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, Guangzhou, 2017

[11]R. M. Mohammad, F. Thabtah, and L. McCluskey, ''An assessment of features related to phishing Websites using an automated technique,'' *In Proc. Int. Conf. Internet Technol. Secured Trans.*, 2012, pp. 492–497.

[12] D. Dua and C. Graff. UCI Machine Learning Repository. School of Information and Computer Science, University of California, Irvine, CA, USA. Accessed: Jan. 10, 2020. [Online]. Available: https://archive. ics.uci.edu/ml/datasets/Phishing+Websites.