International Journal of All Research Education and Scientific Methods (IJARESM),
ISSN: 2455-6211, Volume 11, Issue 5, May-2023, Available online at: www.ijaresm.com

An Intelligent Cyber Security Phishing and Cyber Bullying Detection System Using Machine Learning Approach

Nanditha Balamurali[1], Nisha Basil[2], Shreya H Prabhu[3], T C Sanika[4], Ms. Jismy Mathew[5]

[1,2,3,4]Dept. of CSE, Fisat, Ernakulum, India
[5]Dept. Computer Science, FISAT, Ernakulum, India

----------**********----------

ABSTRACT

Phishing is a form of social engineering attack which affects millions of people every year. With the advancement of technology in recent times, the number of crimes related to phishing has increased exponentially. It has led to a loss of millions of dollars per year. Phishing spreads via e-mail, SMS, instant messaging, social networking, etc. With the increasing number of cases, it is important to provide a layer of protection on the user's side. Cyberbullying, also referred as cyber harassment, is a form of bullying or harassment using electronic means. It is mainly common among teenagers. With the expansion of technology, the number of cyberbullying cases has drastically increased in recent times. Cyberbullying includes threats, rude tweets, aggressive texts, posting pictures, videos or personal information which can harm others. The proposed system aims to address these issues by integrating a phishing website classification and cyberbullying detection system. It seeks to build a highly accurate classification model for identifying phishing websites, emails, and cyber-bullying activities using various machine learning techniques.

INTRODUCTION

Concerns about security issues have become more intense with developments in internet technologies and the consequent revolution in online user interaction. The evolving security issues pose threats to the internet user and may lead to monetary and identity loss for the user.

Phishing is a kind of social engineering threat that exploits the ignorance of uninformed internet users to obtain sensitive information from them in a deceiving manner. Phishers or attackers present themselves as genuine internet users.Phishers try to gain unauthorised access to a victim's accounts in order to steal sensitive or personal information and the victim's identity.

The proposed -system is a combination of phishing website classification and cyberbullying detection. The phishing website classification utilizes a hybrid approach, including blacklist and whitelist, heuristics, and visual similarity, and uses machine learning techniques such as Random Forest, SVM, and Decision Tree. The cyberbullying detection system detects, identifies, and classifies cyberbullying activities from a large volume of live chat texts. These texts are subjected to cluster and discriminant analysis to identify abusive texts. The abusive texts are then grouped using Bidirectional LSTM as the classification algorithm to train datasets and create a predictive model.

LITERATURE SURVEY

A. A Deep Learning-Based Framework for Phishing Website Detection

This paper proposes a deep-learning based framework for detecting phishing websites. It is implemented as a browser plug-in which detects the phishing risk in a web page and alerts the user if there is any risk It provides a real-time environment without any delays and third party services. Four elements make up the proposed framework: data collection activities, machine learning (ML), cloud applications, and web browser extensions. The data collection module is a programme for scheduled tasks. The modules are trained using machine learning modules. False alarm management is done via a cloud-based programme. The data sources used are Phish Storm, Phish Tank, Kaggle and KPT-12. Logistic Regression, RNN, RNN-GRU, LSTM, Support Vector Machine and Random Forest are the six machine learning models used in this framework. The RNN-GRU model with the KPT-12 dataset performs best, with an accuracy of 99.18%, according to the results.

B. DEA-RNN: A Hybrid Deep Learning Approach for Cyberbullying Detection in Twitter Social Media Platform

Social media platforms have seen an increase in the prevalence of cyberbullying (CB). To detect CB on Twitter, a hybrid deep learning model known as DEA-RNN is presented in this paper. For the purpose of fine-tuning the Elman RNN's parameters and reducing training time, the proposed DEARNN model combines an optimized Dolphin Echolocation Algorithm (DEA) with RNNs of the Elman type. Using a dataset of ten thousand tweets, the methodology has evaluated DEARNN and compared its performance to that of cutting-edge algorithms like Bi-LSTM, RNN, SVM, Multinomial Naive Bayes (MNB), and Random Forests (RF). In every scenario, the experimental results demonstrate that DEA-RNN was superior. In terms of detecting CB on the Twitter platform, it performed better than the considered existing methods. DEARNN performed better, achieving an average accuracy of 90.45 percent, precision of 89.52 percent, recall of 88.98 percent, F1 score of 89.25 percent, and specificity of 90.94%. Even though the hybrid proposed model performed better than the other models that were taken into consideration. The current study's sole emphasis was the Twitter dataset. ln addition, the study is restricted to tweets' contents. It was not possible to conduct the behavior-based analysis on the users.ln addition,the objective to classify and locate CB tweets in a live stream has been achieved.

C. Multilayer Stacked Ensemble Learning Model to Detect Phishing Websites

Phishing is a cyber attack that uses a fake website to pretend to be a legitimate one to get users to give out personal information. The attackers can use the stolen credentials to access not only the targeted website but also other popular legitimate websites. Although numerous toolbars, extensions, and anti-phishing techniques exist to combat phishing websites, phishing attacks remain a major issue in today's digital world. The paper proposes a multilayered stacked ensemble learning method with estimators at various layers that feed the predictions from the current layer into the next layer as input. Experimental results show that the proposed model performs well with an accuracy range from 96.79 to 98.90 percentage when compared with different datasets.The findings indicate that balanced data performed better for MLSELM than unbalanced data. Additionally, the proposed model achieved significant differences in a variety of evaluation metrics and outperformed various baseline models.

D. Phishing URL Detection : A Real-Case Scenario Through Login URLs

Phishing is a cyber attack where users are tricked into entering their login details into a website and then transferring that information to a malevolent server using social engineering-An approach that can use URL analysis to identify phishing websites is presented in the paper, which compares and contrasts machine learning and deep learning methods. The legitimate class in the majority of current cutting-edge phishing detection solutions consists of homepages without login forms.However, because it simulates a real-world scenario and because testing current methods with URLs from genuine credentials results in a significant falsified rate. It has been shown that models lose accuracy With time utilising datasets.

Additionally, a frequency analysis has been conducted of the most prevalent phishing domains to identify the various strategies utilized by phishers in their campaigns. A new dataset has been created called Phishing Index Login URL (PILU-90K) to support these assertions. It consists of 60 thousand legitimate URLs, such as index and login websites, and 30 thousand phishing URLs. The advantage of the method used in this paper is that unlike other methods, which use homepage URLs as legitimate class representatives, this model was trained using legitimate login websites.

E. Robust Ensemble Machine Learning Model for Filtering Phishing URLs: Expandable Random Gradient Stacked Voting Classifier (ERG-SVC)

The paper presents a machine learning model for accurately detecting phishing attacks, verified by various datasets. A feature selection approach is employed to create a lightweight pre-processor, and both supervised and unsupervised techniques are used in the detection process. Seven classification algorithms, one clustering algorithm, two ensemble methods, and two large standard legitimate datasets (73,575 and 100,000 URLs) are utilized. The evaluation is performed using two test modes (percentage split and K-Fold cross-validation) and the final predictions are made using a voting classifier ensemble model. The results are compared to those of other methods.

F. Sufficiency of Ensemble Machine Learning Methods for Phishing Website Detection

It compares the performance of various machine learning and deep-learning algorithms. It also explains the reasons why ensemble learning techniques are superior to other models for binary phishing detection. Random Forest shows the better performance when compared to other traditional machine learning algorithms with an accuracy of 97.01 Random Forest also outperforms other ensemble methods in both accuracy and value.

The highest performance in phishing categorization is typically achieved using ensemble machine learning techniques, particularly boosting techniques. The accuracy of deep learning-based approaches like CNN, FCNN, and LSTM is 91.38%, 90.13%, and 89.73%, respectively. According to experimental results, Random Forest outperforms other deep learning models with an accuracy of 96.94%. Random Forest experiences a minimal accuracy deterioration of O. I Ensemble methods reduces the risk of selecting an improper decision. The drawback is that the capability of machine-learning-based systems to manage huge data and extract features is being questioned.

G. Detection of Possible Illicit Messages Using Natural Language Processing and Computer Vision on Twitter and Linked Websites

The paper uses NLP to detect Twitter messages that may support illegal activities and exploit minors. The images and URLs in these messages are analyzed and categorized by gender and age group, with the goal of identifying photos of minors under 14 years old. The process starts by mining tweets with hashtags related to minors. The tweets are then cleaned of noise and misspelled words and classified as suspicious or not.

The models are used to identify the facial features. With the use of SVM and CNN, gender and age group can be recognized based on torso information and its relationship with the head, even when facial details are blurred. The results show that using only torso features with the SVM model results in better performance compared to the CNN model.

COMPARISON

A. Cyber Security Phishing Detection

For cyber security phishing detection, the performances of algorithms such as Support Vector Machine (SVM), Random Forest and Decision Tree is compared based on their strengths and weaknesses for the task.

Table 1: Comparison of Methods

Method	Accuracy
Decision Tree	60.95 %
SVM	73%
Random Forest	60.96%
Bidirectional LSTM	81.852%

SVM is one of the most popular supervised learning algorithms, which is used for classification as well as regression problems. The goal of the SVM algorithm is to create the decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. In the case of phishing detection, SVM shows an accuracy of 73%.

Decision tree uses a information gain measure which indicates how well a given feature separates the training examples according to their target classification. However, decision tree can create over-complex trees that do not generalize the data well. In the case of phishing detection, decision tree shows an accuracy of 60.95%.

Random Forest algorithm is based on ensemble learning, which is a process of combining multiple classifiers to solve a problem and to improve performance of the model. It consists of a number of decision trees on various subsets of the given dataset and takes average to improve predictive accuracy of that dataset. However, random forest can't describe relationships within the data. In the case of phishing detection, random forest shows an accuracy of 60.96%.

B. Cyberbullying Detection

A Bidirectional LSTM, or biLSTM is a sequence processing model that consists of two LSTMs , one taking the input in a forward direction, and the other in a backward direction. BiLSTMs effectively increase the amount of information available to the network, improving the context available to the algorithm. In the case of cyberbullying detection, bidirectional LSTM, shows an accuracy of 81.852%.

Based on the comparison, we have chosen SVM algorithm for implementing phishing detection and Bidirectional LSTM for cyberbullying detection.

METHODOLOGY

A. Dataset

The proposed system includes data from databases like kaggle, fish bank which will be converted to a csv form including all the details required. Then all the necessary data like URLs, path, sub domain, domain, rank, index taken etc will be sorted. All the available data are present in raw form in websites like kaggle and phishtank .The data is then either taken manually or using scraping tools in python and converted to a usable digital CSV format which will act as the primary data that can be used throughout the prediction process.

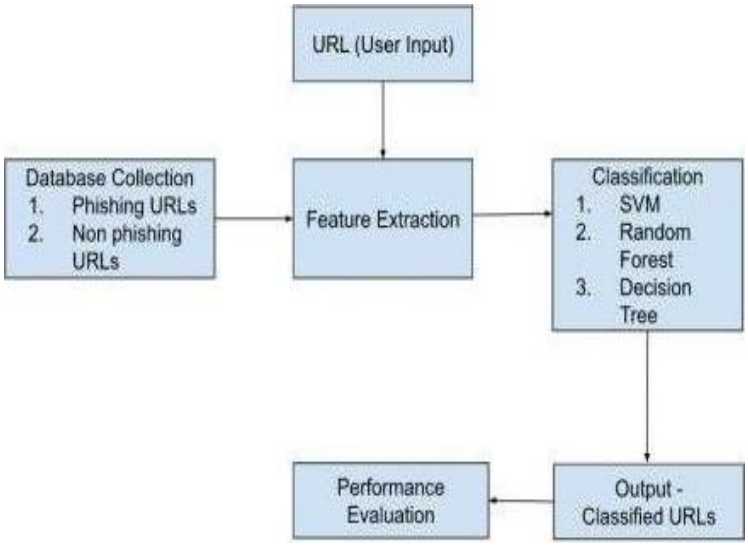

Fig. 1 Work Flow (Phishing)

B. Data Pre-processing

Real-world datasets are typically jumbled, raw, insufficient, inconsistent, and insignificant. It may include manual entry errors, missing values, inconsistent schema, and others. Data pre-processing is the process of transforming raw data into a comprehensible and useful format. To conduct an efficient and accurate analysis, this is a critical stage before implementing any Machine Learning or Data Mining algorithms. Data pre-processing includes various steps such as data cleaning , data integration , data reduction and data transformation. For phishing website classification, features such as the website's URL, domain name, Content and for cyberbullying detection, features such as the text content, author information, metadata can be extracted. Rows with NaN (Not a Number) and None values are removed using the dropna() method.

C. Features Chosen

- Identification and classification of websites as phishing or legitimate.
- Additional layer of protection added on the user side in case of phishing website.
- Classification of tweets from twitter as cyberbullying or not.
- Identification of the type of cyberbullying as racism , sexism or none.
- Bidirectional LSTM for simulating the reciprocal dependencies between words and phrases in the Sequence, both in the forward and backward directions.

The proposed system is basically an integrated model of phishing website classification along with a cyberbullying detection system. The phishing website classification system

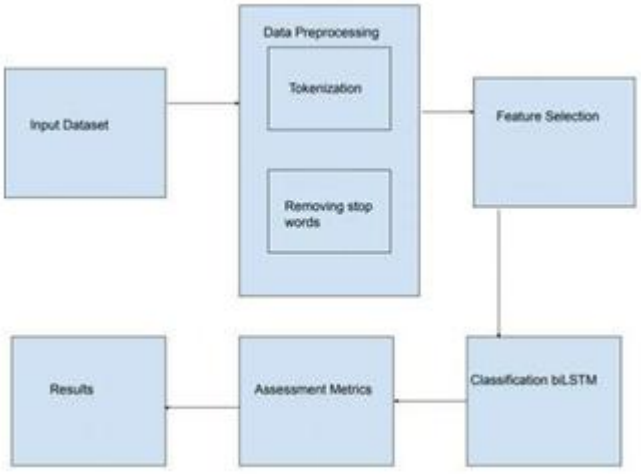

Fig. 2 Work Flow (Cyberbullying)

Uses a hybrid solution based on the following three approaches which are heuristics, visual similarity, blacklist and whitelist. Some machine learning algorithms like Decision Tree, Random Forest and Support Vector Machine will be applied onto collected data. Detection system for cyberbullying will help to detect, identify as well as classify activities from the large volume of streaming texts from Live chatting as cyberbullying or not. Abusive messages will also be identified and classified as racisim , sexism or none. Bidirectional ISTM Algorithm will be used to train the model as it shows maximum accuracy in comparison with other algorithms such as Naive Bayes and Decision Tree. The system would be developed using ML, web technology and python.

CONCLUSION

With the increase in the number of phishing and cyberbullying cases with each passing day, it is essential to protect users. This project aims to provide a safe experience for users by protecting them against phishing and cyberbullying. The project presents a system to automatically detect texts which imply cyberbullying, including different types of cyberbullying, like racism and sexism, as well as detection of phishing URLs. The SVM algorithm has a maximum accuracy of 73% and it will be used for phishing detection. Bidirectional LSTM has a maximum accuracy of 81.825% and it will be used for detecting cyberbullying texts.

REFERENCES

[1]. [1] Detection of Possible Illicit Messages Using Natural Language Processing and Computer Vision on Twitter and Linked Websites - IEEE Access - March 13,2020
[2]. A Deep Learning-Based Framework for Phishing Website Detection IEEE Access - January 6 , 2022
[3]. Phishing URL Detection Real-Cmse Scenario Through Login URLs
[4]. - IEEE Access - April 27, 2022
[5]. MADMAX: Browser - Based Malicious Domain Detection Through Extreme Learning Machine - IEEE Access - June 3 2021
[6]. 15] An Explainable Multi-Modal Hierarchical Attention Model for Developing Phishing Threat Intelligence - IEEE Transactions March/April 2022
[7]. PDGAN: Phishing Detection With Generative Adversarial Networks IEEE Access March 9 2022
[8]. Generalized Outlier Gaussian Mixture Technique Based on Automated Association Features for Simulating and Detecting Web Application Attacks - IEEE Transactions April-June 2021
[9]. Multilayer Stacked Ensemble Learning Model to Detect Phishing Websites - IEEE Access July 2 2022
[10]. Robust Ensemble Machine Learning Model for Filtering Phishing URLs: Expandable Random Gradient Stacked voting Classifier (ERG-SVC)IEEE Access November 2021
[11]. Sufficiency of Ensemble Machine Learning Methods for Phishing Websites Detection - IEEE Access November 30 2022
[12]. A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques - IEEE Access June 2022
[13]. DEA-RNN: A Hybrid Deep Learning Approach for Cyber bullying
[14]. Detection in Twitter Social Media Platform - IEEE Access March 2022 [13] A. S. Srinath, H. Johnson, G. G. Dagher, and M. Long,
[15]. "BullyNet: Unmasking cyber bullies on social networks," IEEE Trans-computed. Social syst., vol. 8, no. 2, pp. 332-344, Apr. 2021, doi:10.1 IOWTCSS.2021.3049232.
[16]. [14] Springer, https://hnk.springer.com/artic1e/10.1007/s10586-022-03604-4