**DTSC5303 CYBERSECURITY IN DATA SCIENCE**

**Professor: Dr. Al Pisano**

**Student: Abhijay Tyagi**

**PHISHING DETECTION**

**Problem Statement:** The project's main objective is to develop a model that can help us to detect email phishing attempts from the text of email bodies using natural language processing and machine learning.\

**How I relate to Phishing Frauds-**

My own experience with a phishing assault has sharpened my awareness of the risks and complexity associated with these types of schemes. One day, I got an email claiming there was unusual activity on my account, supposedly from a reputable online payment company. The email threatened to suspend my account if I didn't click on a link to confirm my account credentials right away.

Nevertheless, a deeper look revealed a few warning signs. The sender's email address sounded strange and didn't match the payment platform's official domain. The email also appeared amateurish overall and contained numerous grammatical problems.

I chose to enter my account directly through the official website rather than clicking on the offered link, and there I saw no signs of any problems or security alarms.

This phishing attempt I encountered made me realize how important it is to deal with unsolicited emails with caution and vigilance. It also acted as a sobering reminder of the need to confirm the legitimacy of email sources, never click on dubious links, and never divulge personal information in response to such communications. One important step in keeping others safe from similar scams was to report the phishing effort to the official organization.

**SOCIAL AND ETHICAL IMPACTS OF PHISHING ATTACKS**-

Phishing assaults impact people, companies, and society at large, with significant social and ethical ramifications. These schemes violate people's privacy by gaining unauthorized access to private and sensitive data, which frequently results in fraud and identity theft. Attacks like this reduce people's confidence in online conversations and increase their skepticism toward genuine communications. Victims could suffer from severe mental distress, severe financial loss, and reputational harm.

Furthermore, phishing is an intrinsically unethical practice that involves theft and deception. Vulnerable people are disproportionately affected by these attacks, which have regulatory and legal ramifications. To safeguard people and uphold confidence in the digital realm, combating phishing goes beyond implementing technological security measures. It is also a crucial social and ethical obligation.

**PHISHING DETECTION CLASSIFICATION DETAILS-**

**Introduction-**

With the increasing prevalence of online contacts in our daily lives, phishing attempts represent a serious threat to people, businesses, and cybersecurity in general. Phishing is a dishonest tactic that coerces people into divulging personal information, frequently resulting in identity theft, monetary loss, and privacy violations. By evaluating and categorizing dubious emails and texts, this project seeks to use machine learning and natural language processing (NLP) to fight phishing attacks. Our initiative aims to improve the capacity to identify and counteract phishing efforts by utilizing advanced text analysis and predictive modeling, ultimately making the internet a safer and more secure place. This initiative is a big step toward accomplishing the goal of developing creative and proactive approaches to phishing detection, which is essential given the constantly changing threat landscape.

Research Directed Questions-

1. What techniques are required to improve the dataset?

Extracting text from the Dataset- Using TfidVectorizer , created a corpus of text which will be splitted into training and testing dataset later on.

Data Cleaning: Address missing values and outliers, as these can negatively impact model performance. You can choose to impute missing values or remove data points.

Labeling Accuracy: Ensure the accuracy of labels in the dataset. Mislabeling can lead to inaccurate model training.

1. Data description- Phishing email dataset from the Association of Computational Linguistics.This dataset includes English email bodies with binary labels indicating whether each message is a real message or a phishing attempt. A dataset of 130424 features and 11928 items was created after the text was converted into numerical feature vectors using the word2vec neural network. The goal of the research was to develop a model that could determine from an email's body whether it was a genuine communication or a phishing effort.
2. How can we minimize false positives and false negatives?

In the field of cybersecurity, reducing false positives and false negatives in a phishing detection model is essential. While false negatives, in which phishing emails are missed, present major security concerns, false positives, in which legitimate emails are inadvertently marked as phishing, have the potential to sabotage communication and undermine confidence. It takes a comprehensive strategy to find a balance. High-quality training data and carefully designed features that distinguish between phishing and authentic emails are the first steps in the process. The model's accuracy can be improved by using ensemble models, adjusting classification thresholds, and utilizing real-time and behavioral analysis techniques. Along with the development of feedback systems for model refinement, continuous learning and user education are essential elements. In the end, the search for a successful phishing detection model depends on a flexible, dynamic, and multidimensional approach that changes with the threat environment.

**Methodology-**

Initially, we loaded data as a pd data frame then we created a corpus of the text from email bodies and labeled it with class 1 or 0 then we dropped all null values and we analyze the dataset using exploratory data analysis (EDA), extracting all of its characteristics (count, mean, standard deviation, min, quantiles, and max).

Thorough data preparation methods are used to clean, prepare, and organize the dataset. This covers managing missing values, encoding characteristics that are categorical, and guaranteeing the accuracy of the data. Some highlights of the dataset-

Data columns (total 2 columns):

Text 11928 non-null object

Class 11928 non-null int64

|  | **Class** |
| --- | --- |
| **count** | 11928.000000 |
| **mean** | 0.434775 |
| **std** | 0.495748 |
| **min** | 0.000000 |
| **25%** | 0.000000 |
| **50%** | 0.000000 |
| **75%** | 1.000000 |
| **max** | 1.000000 |

Once the data preprocessing is completed, we perform data splitting and keep 40% of data for testing and the remaining 60% for training.

A screen shot of a computer code

Description automatically generated

Then we applied KNN and calculated and plotted AUC score for training and testing-

A graph of a number of neighbors

Description automatically generated

A graph with a line

Description automatically generated

Best number of neighbors training: 1

Highest AUC score training: 0.9998392799742848

Best number of neighbors testing: 1

Highest AUC score testing: 0.7771084337349398

A graph of dataset curves

Description automatically generated

**Then we applied SVM –**

**Results after applying SVM:**

Linear kernel training AUC: 0.9999937221844097

Linear kernel testing AUC: 0.9979268352609548

Poly kernel training AUC: 1.0

Poly kernel testing AUC: 0.9986702762105151

Linear kernel training time: 70.862

Linear kernel prediction tim: 7.589

Poly kernel training time: 274.722

Poly kernel prediction tim: 31.623

A graph of dataset curves

Description automatically generated with medium confidence

**Result and Conclusion-**

It is clearly visible that we got better results with SVM rather than KNN algorithm. The major drawback of this model is that it requires heavy computing power which should be taken into consideration for future analysis.

**Real World Examples-**

RSA Security Breach(2011): In one of the worst cyberattacks, attackers gained access to the security company RSA with spear-phishing emails. Sensitive data pertaining to RSA's SecurID tokens—which are frequently used for two-factor authentication—was stolen as a result of the assault.

Sony Pictures Entertainment(2014): A sophisticated phishing technique was used in the Sony Pictures hack, which was linked to North Korea. Sony personnel received fraudulent emails from attackers, which resulted in the disclosure of private company information, employee information, and unreleased film.

John Podesta's Email (2016): John Podesta, the chairman of Hillary Clinton's campaign, was the target of a phishing attempt in the course of the 2016 US presidential election. Due to the hacking of his personal email account, WikiLeaks was able to obtain confidential campaign data.

Google docs Phishing (2017): 2017 saw the global phishing assault known as Google Docs Phishing, which invited Gmail users to open a Google Docs document. Users' email account credentials were stolen as a result of the malicious link's redirection to a phony Google login page.

**Resources-**

* <https://scholar.smu.edu/cgi/viewcontent.cgi?article=1215&context=datasciencereview>
* <https://www.sciencedirect.com/science/article/pii/S1877050921011741>
* <https://www.mdpi.com/2076-3417/13/9/5275>
* <https://ieeexplore.ieee.org/document/9795286>
* <https://www.ncsc.gov.uk/guidance/phishing>
* <https://www.ibm.com/topics/phishing>
* <https://www.loginradius.com/blog/identity/real-time-techniques-detect-phishing-attacks/#:~:text=Email%20and%20Content%20Analysis,links%2C%20to%20identify%20potential%20threats>.
* <https://ieeexplore.ieee.org/document/4149954>