“There are some [phishing Web sites] that are going to be there for two weeks and others that will be there for a few hours. Signature-based antivirus is totally at the mercy of when [the virus] is identified to when it is inoculated against -- that is a huge issue. With the constant threat of zero-day attacks, signatures can't carry us forward,” ~ James Turner [5]

With the increase in online transactions, it becomes very easy to attack credulous people by making them disclose the personal information (PII) like usernames, passwords, card details, OTPs etc. by using counterfeit URLs, aka Phishing attacks. 1 in 99 emails are phishing attacks. 25% of phishing emails manage to make their way into Office 365 inboxes. (cybertak.org newsletter 2022)[6]

Phishing attacks are the practice of sending fraudulent communications that appear to come from a reputable source. Phishing can be conducted via a text message, social media, or by phone, but the term 'phishing' is mainly used to describe attacks that arrive by email. Around 156 million phishing emails are sent every day. Among these 16 million phishing emails make it through filters. Phishing emails can reach millions of users directly and hide amongst the huge number of benign emails that busy users receive. 8 million phishing emails are opened by the users, 800,000 links are clicked from the phishing emails and 7% have replied to spoof or phishing mail unknowingly. 80,000 people fall for the scam every day and give away personal information to people who are trying to phish them. Attacks can install malware (such as ransomware), sabotage systems, or steal intellectual property and money. The goal is to steal sensitive data like credit card and login information. Phishing is a common type of cyber-attack that everyone should learn about to protect themselves.[5]

Phishing emails can hit an organisation of any size and type. People might get caught up in a mass campaign (where the attacker is just looking to collect some new passwords or make some easy money), or it could be the first step in a targeted attack against your company, where the aim could be something much more specific, like the theft of sensitive data. In a targeted campaign, the attacker may use information about your employees or company to make their messages even more persuasive and realistic. This is usually referred to as spear phishing. 95% of phishing e-mails pretend to be from Amazon, eBay, or banks.

* Sources of Phishing:

1. The means of distribution of phishing URLs include, among others, spam, or phishing messages with links to the phishing site, Blackhat search engine optimization (SEO) techniques, Internet downloads, peer-to-peer (P2P) file sharing networks, social networking sites, visiting vulnerable web sites such as blogs, forums, comment accepting news portals, instant messaging (IM), Internet Relay Chat (IRC), etc.
2. The scammers try to mask their IP addresses as well as try shortening the URLs using various shortening services available like tiny URL, bit.ly, etc and trick or mislead the users into clicking into various malicious links and redirect them into Phishing websites. These malicious sites are built with the aim of stealing valuable information and passwords.[1]

* Types of phishing attacks

1. Deceptive phishing: Deceptive phishing is the most common type of phishing. In this case, an attacker attempts to obtain confidential information from the victims. Attackers use the information to steal money or to launch other attacks. A fake email from a bank asking you to click a link and verify your account details is an example of deceptive phishing. Boost your email security (1:49 min)
2. Spear phishing: Spear phishing targets specific individuals instead of a wide group of people. Attackers often research their victims on social media and other sites. That way, they can customize their communications and appear more authentic. Spear phishing is often the first step used to penetrate a company’s defences and carry out a targeted attack.
3. Whaling: When attackers go after a “big fish” like a CEO, it’s called whaling. These attackers often spend considerable time profiling the target to find the opportune moment and means of stealing login credentials. Whaling is of particular concern because high-level executives can access a great deal of company information.
4. Pharming: Like phishing, pharming sends users to a fraudulent website that appears to be legitimate. However, in this case, victims do not even have to click a malicious link to be taken to the bogus site. Attackers can infect either the user’s computer or the website’s DNS server and redirect the user to a fake site even if the correct URL is typed in. [3]

* Purpose of the study:

In our paper, we are going to demonstrate our tool, which not only classifies a given URL as Phishing or non-phishing but also extracts the URL from a text message, even if it is hidden as a hyperlink or in a shortened form, and then using Machine Learning model, the probability of Phishing in an URL can be predicted.

Mainly the financial institutes like banks face these types of attacks. So, they suffer from Reputational damage, Customer churn, overall loss of Brand value. For that Banks and Financial institutions are now seeking ways to prevent Phishing attacks to make their customers feel Safe and Secure.

* Objective of the study:

The main objective of the study is to classify an URL as Phishing or non-Phishing based on the propensity of Phishing involved. The anti-Phishing tool, that we are going to demonstrate will analyse an URL and use feature extraction technique to identify the various lexical and keyword-based features. Thereafter using various ML techniques, we will build a model using which the probability of Phishing in an URL can be predicted.

The Anti-phishing tool classifies URLs in three categories – Malicious, Suspicious and Genuine.

Phishing Url: Whittaker et al. define a phishing web page as “any web page that, without permission, alleges to act on behalf of a third party with the intention of confusing viewers into performing an action with which the viewers would only trust a true agent of the third party.” This definition, which is like the definition of “web forgery”, covers a wide range of phishing pages from typical ones – displaying graphics relating to a financial company and requesting a viewer’s personal credentials – to sites which claim to be able to perform actions through a third party once provided with the viewer's login credentials. Thus, a phishing URL is a URL that leads user to a phishing web page.[1]

Malicious url: A malicious URL is a clickable link that directs users to a malicious or otherwise fraudulent web page or website.

Suspicious URL: A suspicious URL is a URL which appears suspicious but might not be malicious. It appears malicious due to the domain name. Scammers will doctor domains to make them seem like something they aren’t. In the example of http://google.com.cust\_login.ie, the domain is cust\_login.ie, not google.com. And in http://accounts\_login.cz/google.com, the domain is accounts\_login.cz, not google.com.

Characteristics of a Malicious or suspicious url:

The email received promises unrealistic threats or demands. These types of emails eventually ask people to send money or share their login credentials.

One of the most common characteristics of these emails are that they are usually littered with spelling mistakes and poor grammatical errors.

They mostly contain a dodgy URL, and the domain name mostly will contain digits.

These type of Phishing URLs can be identified through the lexical based and keyword based features.

References:

1. Basnet, Ram. (2014). LEARNING TO DETECT PHISHING URLS. International Journal of Research in Engineering and Technology. 03. 11-24. 10.15623/ijret.2014.0306003.
2. <https://www.milnsbridge.com.au/5-characteristics-phishing-email/>
3. <https://www.cisco.com/c/en_in/products/security/email-security/what-is-phishing.html>
4. Ma, Justin & Saul, Lawrence & Savage, Stefan & Voelker, Geoffrey. (2009). Beyond blacklists: learning to detect malicious Web sites from suspicious URLs. 1245-1254. 10.1145/1557019.1557153.
5. <https://sites.google.com/site/phishingsimplified3/phishing-quotes-and-statistics>
6. <https://www.cybertalk.org/2022/03/30/top-15-phishing-attack-statistics-and-they-might-scare-you/>