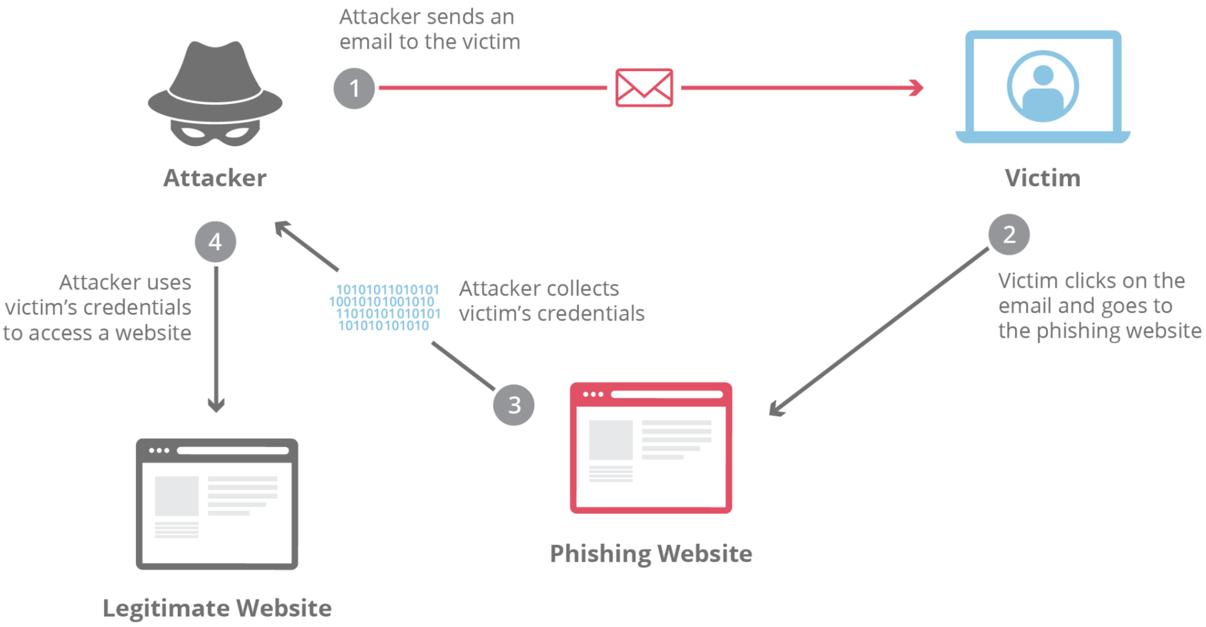
What is Phishing?

- Phishing is a type of cyber attack in which an attacker disguises themselves as a trustworthy entity in order to trick the victim into sharing sensitive information such as passwords, credit card details, or other personal data.
- Phishing attacks typically involve the use of fraudulent emails, text messages, or websites that look like they come from a legitimate source such as a bank, online retailer, or social media platform.
- The **goal of phishing** is to steal sensitive information or install malware on the victim's device, which can be used to carry out further attacks or to gain access to the victim's accounts.
- It is important to be cautious when opening emails or clicking on links from unknown sources and to verify the authenticity of any request for sensitive information before providing it.


Attacker sends an email to the victim
1
Attacker
Victim
2
Victim clicks on the email and goes to the phishing website
4
Attacker uses victim's credentials to access a website
10101011010101
10010101001010
11010101010101
101010101010
Attacker collects victim's credentials
3
Phishing Website
Legitimate Website

How can someone detect the phishing?

Phishing attacks can be difficult to detect because they often appear to come from a legitimate source and use convincing language and imagery to trick the victim. However, there are several ways to spot a phishing attempt:

- **Check the sender's email address**: Phishing emails often use fake email addresses or spoofed email addresses that appear to be from a legitimate source. Check the sender's email address carefully to ensure that it matches the official email address of the organization.

➔ **Look for spelling and grammar errors**: Phishing emails often contain spelling and grammar errors, which can be a red flag that the message is not from a legitimate source.

➔ **Look for the lock icon**: If the website requires you to enter sensitive information, such as passwords or credit card details, make sure the website has a lock icon in the address bar. This indicates that the site is using a secure connection.

➔ **Look for contact information**: Legitimate websites usually have contact information such as a phone number or email address. If the website does not have any contact information or the contact information provided does not seem legitimate, this could be a sign that the website is fraudulent.

➔ **Verify the link before clicking**: Hover over any links in the email or message to see the URL before clicking. If the link is different from the official website of the organization or looks suspicious, do not click on it.

➔ **Check for urgency or threats**: Phishing emails often create a sense of urgency or use threatening language to pressure the victim into taking action. Be wary of any email that demands immediate action or threatens negative consequences.

➔ **Use anti-phishing software**: Anti-phishing software can help detect and block phishing emails and websites, and can alert you if you are visiting a suspicious site.

If you suspect that you have received a phishing email, do not click on any links or provide any personal information. Instead, report the email to the organisation it appears to be from and delete it from your inbox.

What is the scope of AI for detecting phishing in a website?

The scope of Artificial Intelligence (AI) for detecting phishing in a website is significant. With the increasing frequency and sophistication of phishing attacks, AI has become an essential tool in the fight against cybercrime. AI algorithms can analyze large amounts of data and identify patterns that are indicative of phishing attacks.

Some specific ways in which AI can be used for detecting phishing in a website include:

➔ **Machine learning algorithms** can be used to analyze website content, URLs, and user behavior to identify patterns that are indicative of phishing attacks.

➔ **Natural language processing (NLP)** can be used to analyze text and identify phishing email messages or other types of content that are designed to deceive users.

➔ **Image recognition** can be used to analyze website images and identify logos or other visual cues that are associated with phishing attacks.

- ➔ **User behavior analysis** can be used to identify patterns of behavior that are consistent with phishing attacks, such as clicking on suspicious links or entering sensitive information on a website.
- ➔ **AI-powered chatbots** can be used to educate users on the risks of phishing attacks and how to identify them.

Overall, AI has significant potential for detecting phishing in a website and can help to reduce the risk of fraud and data breaches. By leveraging machine learning, NLP, image recognition, and user behavior analysis, organizations can build sophisticated systems that are capable of detecting phishing attacks with a high degree of accuracy.

How can ML help us in detecting phishing in a website?

Machine learning (ML) algorithms can help detect phishing in a website by analyzing patterns and characteristics of phishing attacks. Here are a few ways in which ML can be used for this purpose:

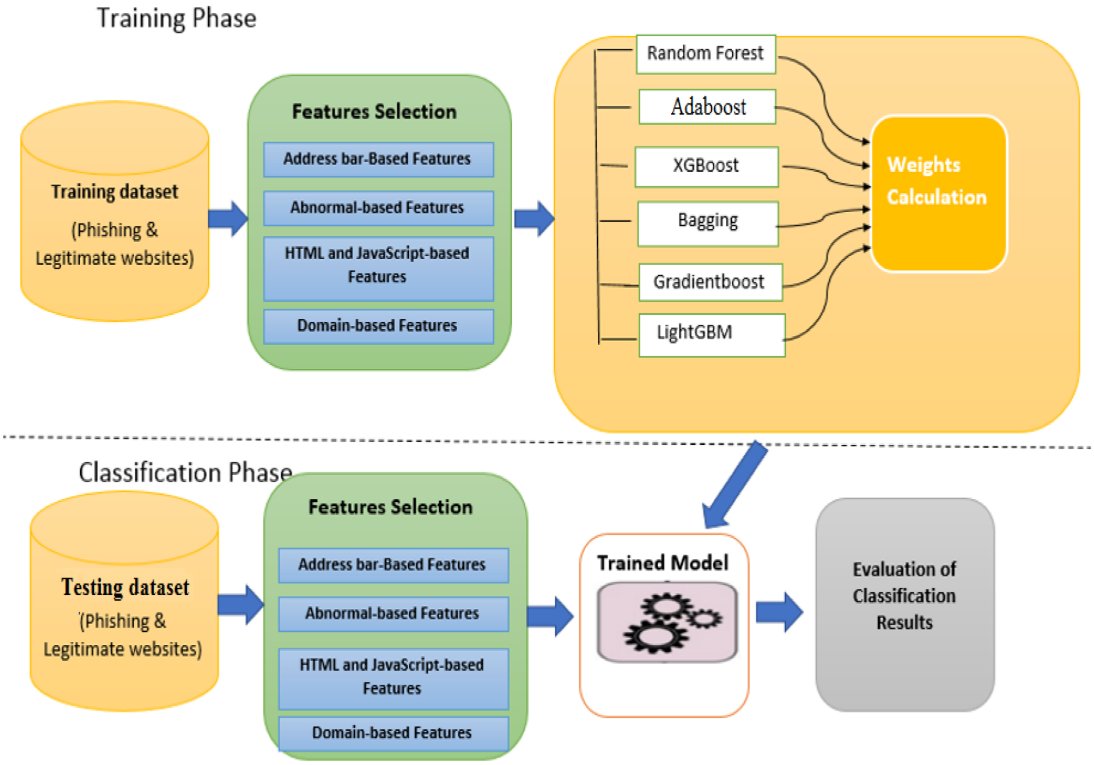
- ➔ **Training on a large dataset**: ML algorithms can be trained on large datasets of known phishing attacks to recognize patterns and features that are common to these attacks. These features can include the URL structure, website content, and user behavior. By analyzing these features, the ML algorithm can learn to identify phishing attacks with a high degree of accuracy.
- ➔ **Anomaly detection**: ML algorithms can be used to identify anomalies in website traffic or user behavior that may be indicative of a phishing attack. For example, if a user is clicking on suspicious links or entering sensitive information on a website, the ML algorithm can flag this behavior as potentially indicative of a phishing attack.
- ➔ **Real-time analysis**: ML algorithms can be used to analyze website traffic and user behavior in real-time, allowing for the rapid detection of phishing attacks as they occur.
- ➔ **Adaptation to new attacks**: ML algorithms can be trained to adapt to new types of phishing attacks by analyzing and learning from new data. This allows for the continuous improvement of the system's ability to detect and prevent phishing attacks.

Overall, ML can be a powerful tool for detecting phishing in a website by analyzing patterns and identifying anomalies in website traffic and user behavior. By leveraging ML algorithms, organizations can reduce the risk of fraud and data breaches caused by phishing attacks.

Procedure for detecting Phishing using ML

Here's a step-by-step procedure for detecting phishing in a website using machine learning:

- ➔ **Collect training data**: The first step is to collect a large dataset of known phishing attacks. This dataset should include features such as the URL structure, website content, and user behavior.
- ➔ **Preprocess the data**: Once you have collected the data, you need to preprocess it to extract relevant features and prepare it for analysis. This may involve data cleaning, normalization, and feature engineering.
- ➔ **Train the machine learning model**: Using the preprocessed data, you can train a machine learning model to recognize patterns and features that are common to phishing attacks. This may involve using supervised learning algorithms such as logistic regression, decision trees, or neural networks.
- ➔ **Test the machine learning model**: Once you have trained the model, you need to test it on a separate dataset of known phishing attacks. This will help you evaluate the accuracy and effectiveness of the model.
- ➔ **Deploy the model**: Once you have tested the model and are satisfied with its performance, you can deploy it in a production environment to detect phishing attacks in real-time. This may involve integrating the model with existing security systems or using APIs to access the model's functionality.

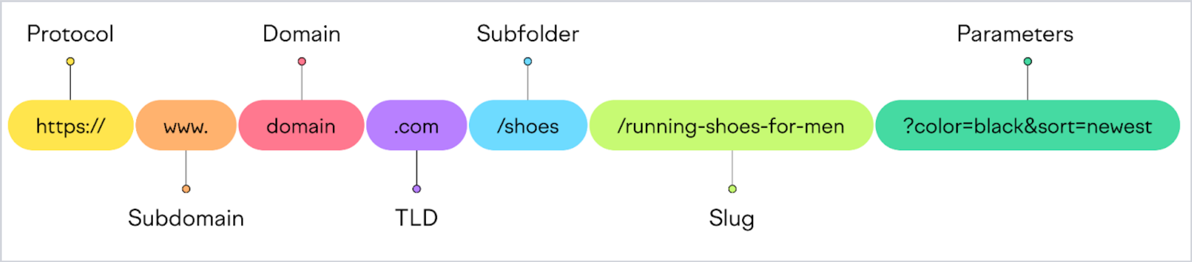


- ➔ **Monitor and update the model**: It's important to monitor the model's performance over time and update it as new types of phishing attacks emerge. This may involve retraining the model on new data or adjusting its parameters to improve its accuracy.

Overall, the key to success in detecting phishing using machine learning is to have a large dataset of high-quality training data and to continually monitor and update the model as new types of phishing attacks emerge. With the right approach, machine learning can be a powerful tool for detecting and preventing phishing attacks in a website.

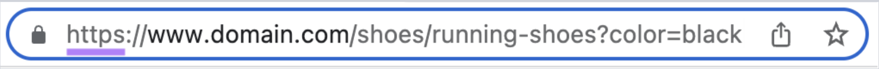
Uniform Resource Locator (URL)

The structure of a URL breaks down into seven distinct parts. Like this:



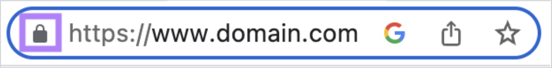
1. Protocol

- ➔ The protocol tells your browser how to connect to a webpage.



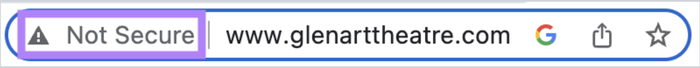
https://www.domain.com/shoes/running-shoes?color=black

- ➔ It could be HTTP (hypertext transfer protocol) or HTTPS (HTTP secure).
- ➔ The main difference between the two is that HTTPS encrypts and protects any data transmitted between the server and browser.
- ➔ So when users interact with your site, their sensitive information (like login and credit card details) is safe from attackers.
- ➔ Websites that show a lock icon in the address bar are using HTTPS.



https://www.domain.com

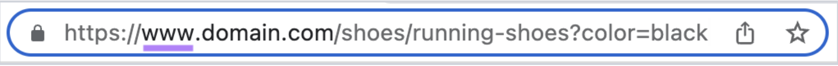
- ➔ Whereas websites that show a "Not Secure" warning in the address bar are using HTTP.



Not Secure | www.glenarttheatre.com

2. Subdomain

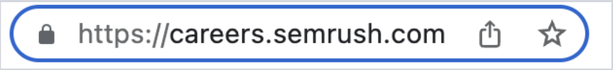
- ➔ A subdomain is a string of letters or a complete word that appears before a URL's first dot.



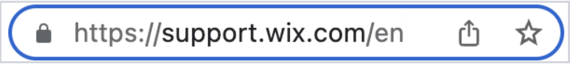
https://www.domain.com/shoes/running-shoes?color=black

- ➔ The most popular subdomain is www. It stands for World Wide Web, communicating that the URL is a web address.

- ➔ In the past, it was common to use www. But nowadays you can omit it from your URLs if you want.
- ➔ It doesn't matter whether you use it or not. It all depends on your personal preference.
- ➔ Then there are other subdomains—blog, store, support, news, careers, and so on—that are used for managing and organizing sections of a website that serve a specific function.
- ➔ For example, we use the "careers" subdomain to organize and display all the career opportunities at Semrush.

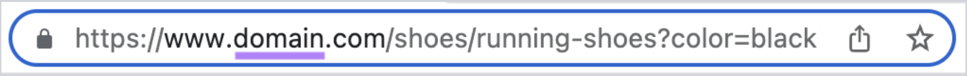

https://careers.semrush.com

- ➔ And Wix uses the "support" subdomain to feature all their help articles on the website.


https://support.wix.com/en

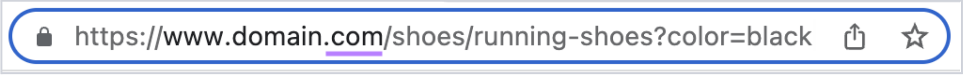
3. Domain

- ➔ A domain is the main part of the URL that identifies the website. Like eBay, Expedia, or Semrush.


https://www.domain.com/shoes/running-shoes?color=black

- ➔ If you're shopping around for a domain name, our advice is to choose something short and catchy.
- ➔ Short and catchy domain names are easier for people to remember. So they are more likely to return to your website.
- ➔ Plus, they're easier to fit into logos and other branding materials compared to long, complicated ones.

4. Top-Level Domain (TLD)


https://www.domain.com/shoes/running-shoes?color=black

- ➔ The TLD (also called domain extension) is the part that comes after the name of your website, like "**.com**."
- ➔ You'll come across many TLDs on the internet. Here is a list of five of the most common ones. And what kinds of websites they're suited for:
  - ◆ **com**: Commercial websites
  - ◆ **org**: Nonprofit organizations
  - ◆ **net**: Software and hosting companies providing network services
  - ◆ **edu**: Educational institutions (universities, colleges, schools, etc.)
  - ◆ **gov**: Government agencies and departments
- ➔ If you're in the early stages of buying a domain name, we recommend choosing a TLD that best describes the nature of your business.

country-code top-level domains(cc TLDs)

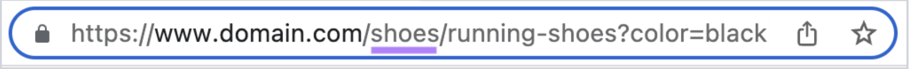
- ➔ Additionally, you can use country-code top-level domains (ccTLDs).

- ➔ ccTLDs are two-letter domain extensions that indicate a website’s association with a specific country or territory.

  Examples include:
  - ◆ in for India
  - ◆ uk for United Kingdom
  - ◆ de for Germany
  - ◆ cn for China
  - ◆ ca for Canada
  - ◆ es for Spain
  - ◆ au for Australia
  - ◆ Using ccTLDs makes sense when a website’s target audience is predominantly based in a specific country.
  - ◆ By using ccTLD, a website signals its connection to that location. Which can help to establish trust and credibility with users in that region.

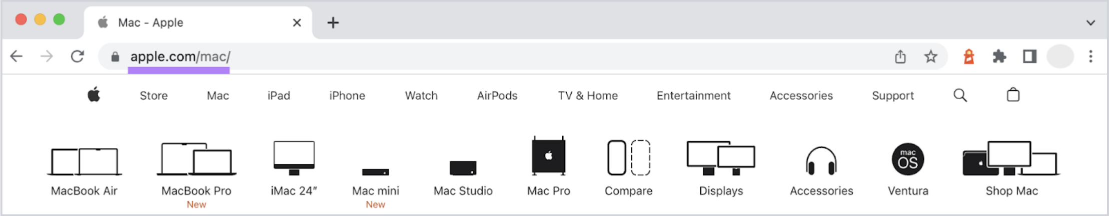
5. Subfolder

- ➔ A subfolder is a folder or directory that is located within the top-most directory (or main directory) in your site hierarchy.



https://www.domain.com/shoes/running-shoes?color=black

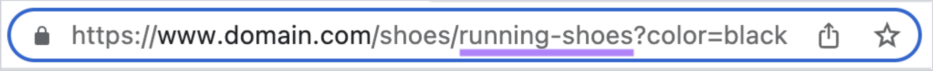
- ➔ For example, consider the URL “www.domain.com/shoes/”
- ➔ In this URL, the “shoes” subfolder is located within the main directory of the website, which is “www.domain.com”
- ➔ Similar to subdomains, subfolders are used to separate website content into logical sections.
- ➔ This makes it easier for visitors to understand where they are on the website.
- ➔ For example, users on this URL “www.apple.com/mac/" can easily understand that they’re on the MacBook section on Apple’s website.



apple.com/mac/

6. Slug

- ➔ A slug is the part of a URL that identifies a specific page or a post on a website. It helps users understand the context and content of a page.

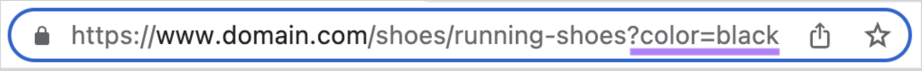


https://www.domain.com/shoes/running-shoes?color=black

- ➔ Look at this URL slug, for example: “/best-baby-shampoos/”
- ➔ Reading this slug alone, users can get an idea of what the page is about.
- ➔ But sometimes, you’ll also come across URL slugs that read like this: “/785321/”
- ➔ Slugs with numbers are confusing to users.
- ➔ The primary goal of the URL slug is to describe the content of a page. So when you create a slug, make sure it’s descriptive.

7. URL Parameters

- URL parameters (or query strings) are part of a URL that comes after a question mark (?).

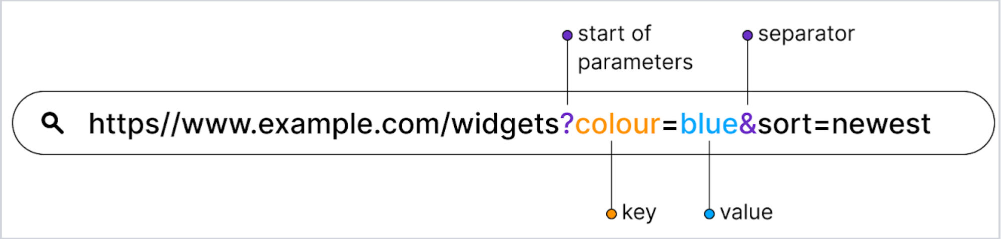


https://www.domain.com/shoes/running-shoes?color=black

- They're composed of keys and values, separated by an equal sign (=).
- The **key** tells you what kind of information is being passed. The **value** is the actual information being passed.

Let's look at an example.

- In the URL below, "color" is the key and "blue" is the value. This parameter will apply a filter to a webpage to display only blue products.
- You can add multiple parameters to a URL by separating them with an ampersand (&).



https//www.example.com/widgets?colour=blue&sort=newest

- Now, there are two parameters: "color" with the value "blue" & "sort" with the value "newest"
- This applies a filter to a webpage to show blue products and sorts them by the newest first.

Parameters serve multiple use cases:

- **Searching** parameters allow you to search results from a website's internal search engine
- **Filtering** parameters let you sort and filter listings on category pages. Listings will often be products, jobs, hotels, flights, etc. And they can be filtered by various attributes, like price, availability, size, brand, salary, location, flight time, delivery time, ratings, etc.
- **Tracking** parameters help you track traffic from your ads and marketing campaigns
- **Paginating** parameters are helpful in organising blog archive pages and forum threads in a series of pages