# A

Phishing Website Detection is the process of identifying phishing websites, which are fraudulent websites that mimic legitimate ones in order to steal sensitive information such as login credentials or credit card numbers. The purpose of my model is to classify a website as either phishing or legitimate based on certain features of the website.

To accomplish this, the model is trained on a dataset of labeled websites, where each website is represented by a set of features such as the URL, the website's content, and the website's structure. Once trained, the model can take in the feature set of a new website and predict whether it is phishing or legitimate.

# B

Phishing is a type of online fraud that involves tricking people into providing sensitive information, such as login credentials or credit card numbers, by pretending to be a trustworthy entity. Phishing websites, also known as spoofed websites, are typically used as part of this scam. These websites mimic the design and branding of legitimate websites, such as those of banks or online retailers, in order to trick people into thinking they are the real thing.

Phishing is a serious problem, with millions of people falling victim to phishing scams each year. The financial loss from phishing is estimated to be in the billions of dollars annually.

There are several theories about why phishing is so effective. One theory is that phishing takes advantage of the human tendency to trust familiar branding and logos. Another theory is that phishing takes advantage of people's trust in their friends and acquaintances, as phishing emails and messages often appear to be from someone the victim knows.

There are several approaches to detect phishing websites, the first one is rule-based systems that are based on predefined rules and heuristics, these systems are easy to implement but they can be easily bypassed by phishers. The second one is based on machine learning, these systems are able to learn from examples and can adapt to new phishing patterns, but they require a large labeled dataset to be trained. Another approach is based on the use of browser extensions, these systems can block phishing websites when they are accessed by the user but they are not able to detect new phishing websites.

The most recent and promising approach is the use of deep learning algorithms, these systems can learn from features of the website such as the website's content, structure, and the URL, these systems are able to detect new phishing websites and have high accuracy rates, but they require a large dataset to be trained.

Overall, phishing website detection is a challenging problem, as phishers are constantly coming up with new ways to evade detection. However, by understanding the theories behind why phishing is effective and using a combination of different approaches, it is possible to develop models that can effectively detect phishing websites.

# C

## k-nearest neighbors (KNN)

1. select only the first two columns of the data (This is used for visualization purposes)
2. plots a scatter plot of the data where each point in the plot represents a sample of the data and its position on the x and y axis corresponds to the values of the first and second columns of the data The color of the points in the scatter plot corresponds to the labels of the samples.
3. Defining a new point and plots it in the same scatter plot
4. Calculating the distance between the new point and all the points in the data. These distances are used to find the k closest points to the new point. The labels of these closest points are used to predict the label of the new point.
5. The code also plots the new points with a color that corresponds to their predicted label.