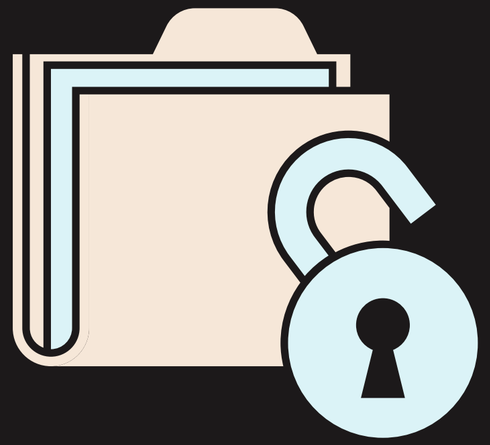
SMARTER SECURITY: PREDICTING PHISHING ATTACKS

Think Before You Click!

Presented by: Gloria Ngure

OUTLINE



1. Overview
2. Business Understanding
3. Data Understanding
4. Modelling
5. Evaluation
6. Recommendations
7. Next Steps

OVERVIEW

- This project aims to leverage machine learning to help proactively identify and block phishing threats before they can cause any harm.

- By the end of this presentation, we will see how a machine learning model can significantly improve security measures and protect both the business and customers.

BUSINESS UNDERSTANDING

- Phishing attacks have become a major concern worldwide. These attacks are getting more sophisticated and the consequences: financial losses, reputational damage, and loss of customer trust: can be severe.

- A robust way to detect these threats early on is needed. The goal with this project is to build a predictive model that can accurately spot phishing websites and help mitigate these risks before they escalate.

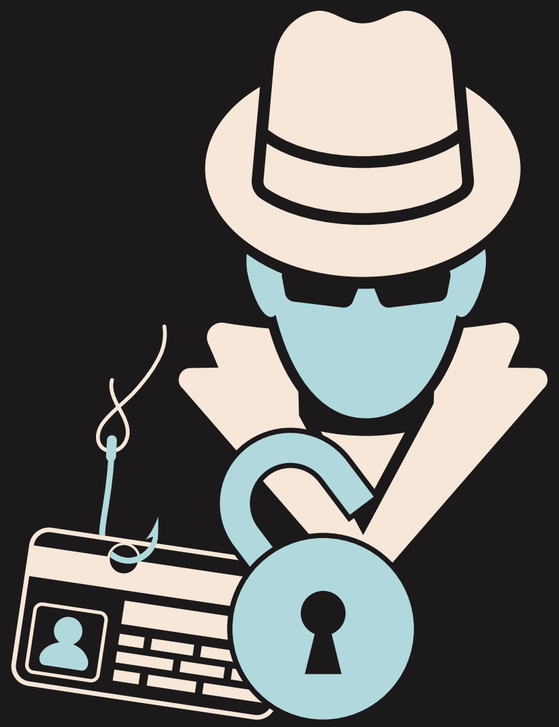
Think of the last time business information was compromised online. How much time, money, and effort would it have saved if we could have prevented that phishing attack from happening in the first place?

DATA UNDERSTANDING

- Two datasets were explored to see which one would best fit the solution. Both datasets were published this year (2024) which means the datasets capture the rapidly evolving phishing tactics used by attackers.

- To build this model, analysis of a variety of website features was conducted. Factors like the length of the URL were looked at — phishing sites tend to have unusually long URLs.

- By understanding the features, a model to distinguish between safe and phishing websites was achieved.



MODELLING

- A method called classification modeling was used for the prediction. Classification is a way for our model to sort websites into two categories: legitimate or phishing.

- This approach is particularly effective because it automates detection, allowing quick and efficient flagging of suspicious websites.

- Two models, Logistic Regression and Decision Tree were employed to see which one performs the best with the given set of inputs.

EVALUATION

- The best classification algorithm was the Logistic Regression Model because it correctly identifies whether a website is legitimate or phishing about 99% of the time. It superseded the Decision Tree which correctly classifies 95% of the time.

- It also had a precision of 99.7% which means, a given prediction is class 1, there is about a 99.7% chance that the model will correctly label it as class 1 (phishing) and about a 0.3% chance that the model will incorrectly label it as class 0 (legitimate).

- For the business, this translates to better protection against phishing attacks, which ultimately safeguards financial assets, reputation, and the trust of customers.

RECOMMENDATIONS

1

Move forward with integrating the logistic regression model into existing security systems.

2

Continuously monitor and update the model to keep it effective against new threats.

3

Conducting user education on identifying phishing attempts and practicing good cybersecurity hygiene.

NEXT STEPS

Train the team on how to use the model and interpret its predictions.

Explore expanding the model’s capabilities to detect other types of online threats, further strengthening the security posture.

THINK BEFORE YOU CLICK!

THANK YOU!

Presented by Gloria Ngure.