APWG

www.apwg.org

Web Phishing Detector

By Peter Vuong

Introduction

Peter Vuong
GitHub
LinkedIn

Agenda

Business Understanding

Data

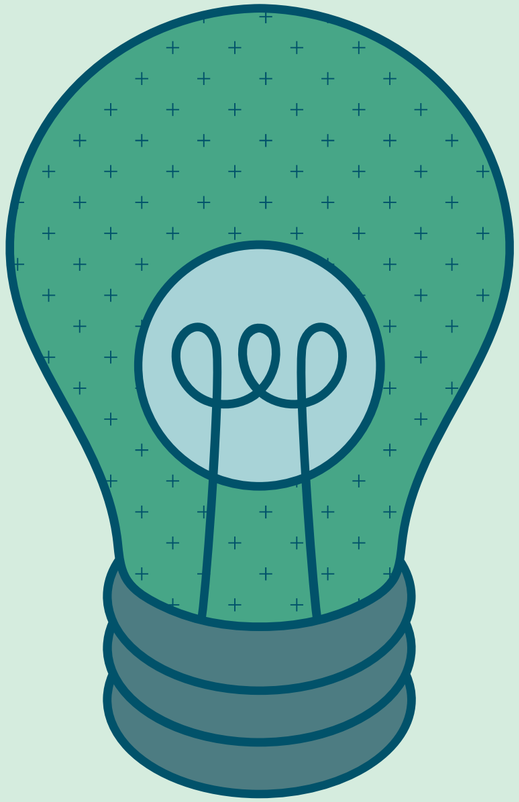
Modeling & Evaluation

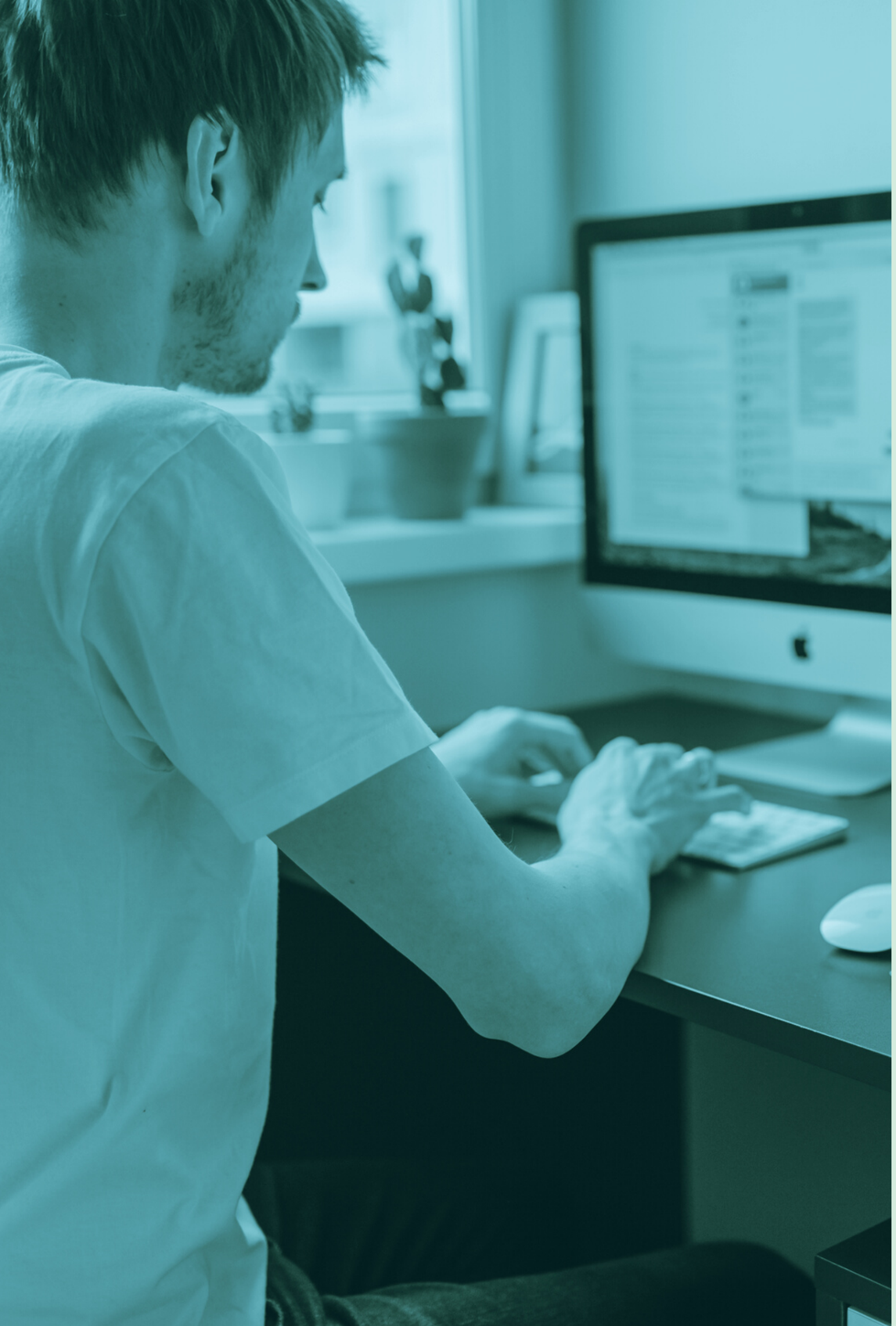
Proposals

Future Considerations

“UNIFYING THE GLOBAL RESPONSE TO CYBERCRIME THROUGH DATA EXCHANGE, RESEARCH AND PROMOTING PUBLIC AWARENESS."

- APWG Mission Statement

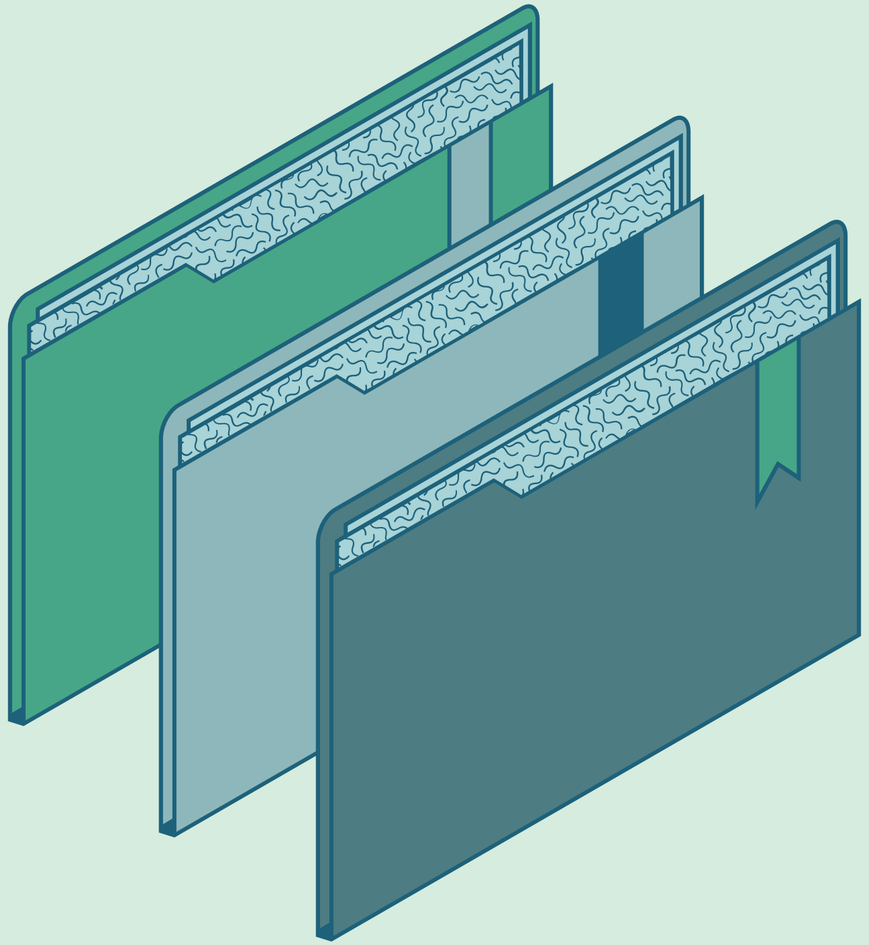




What is phishing?

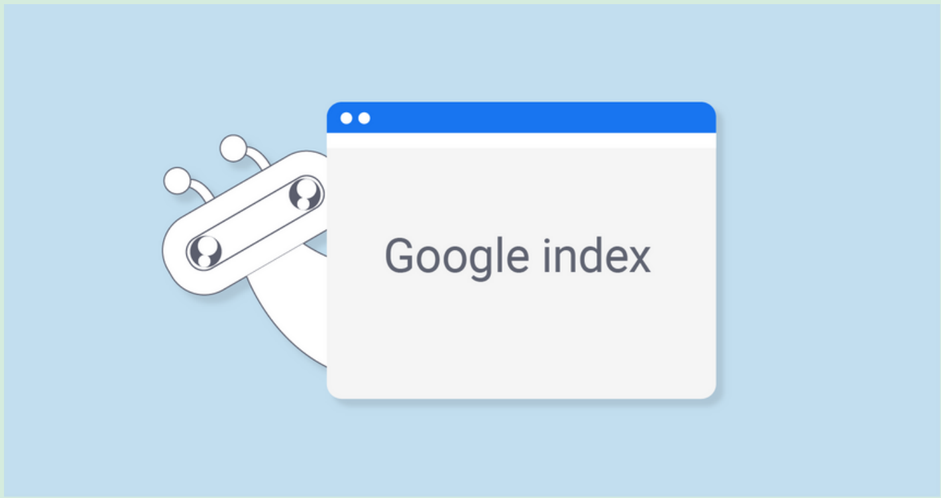
A socially engineered cyber attack

- Commonly aimed towards older people
- Aims to forcibly gain access to sensitive information
- Over 200,000 victims of phishing attacks in the US in 2020



Bottom Line Up Front

- Final model has an accuracy score of 92%
- Recall score of 93%

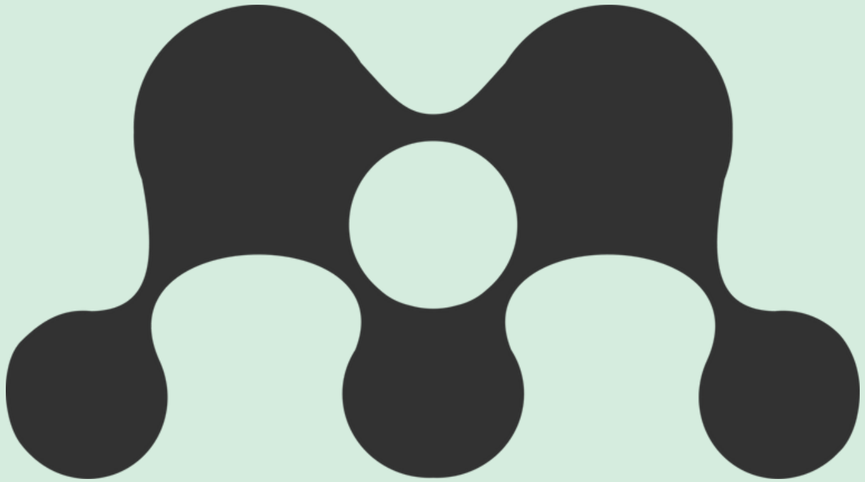


Websites that lack a Google Index are 20x more likely to be a phishing website

Similarly, websites that have phishing hints in their URL are 5x more likely to be a phishing website

includes
signin content
admin plugins view
myaccount
alibaba dropbox site
themes images
login css

Data

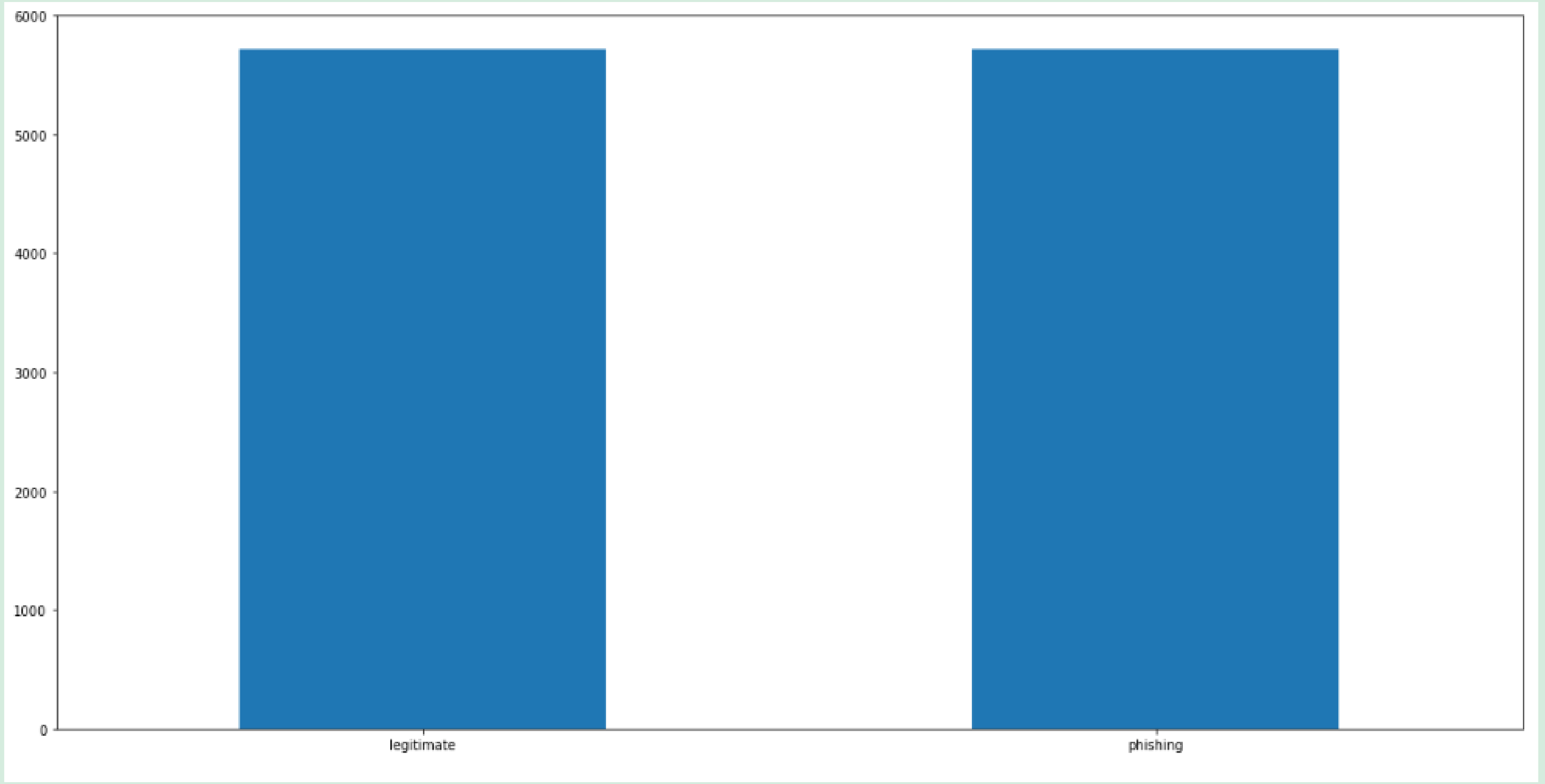


Dataset from Mendeley Data

Published Jun 25, 2021

- 11,430 rows of data
- 87 unique features
- Data from 2020

Distribution of Phishing vs. Legitimate Data



Modeling & Evaluation

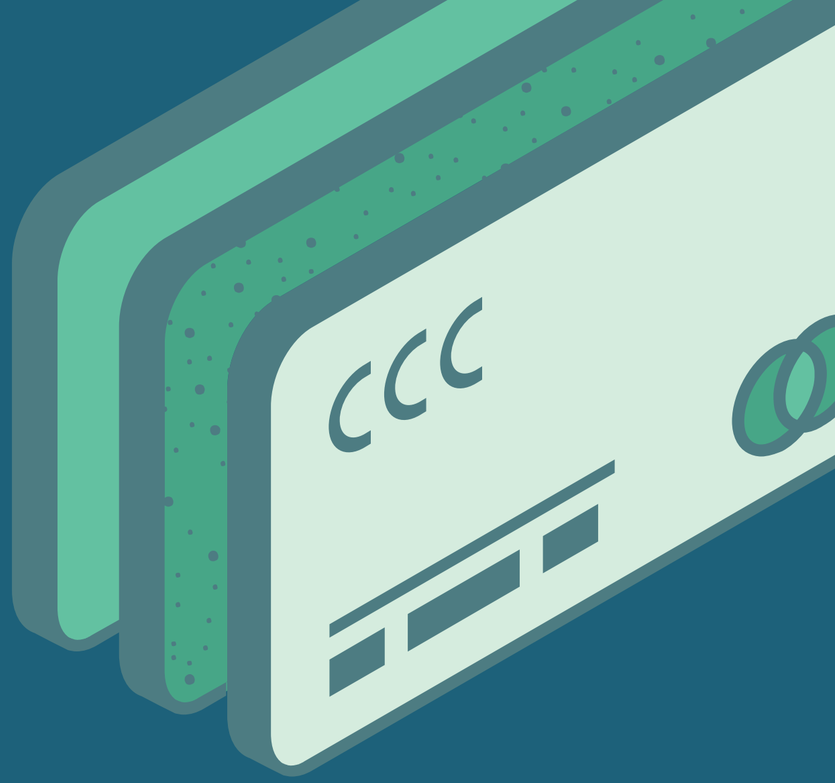
Logistic Regression

Preprocessing

Logistic Regression with ExtraTrees Classifier

Phishing Detection

Logistic Regression

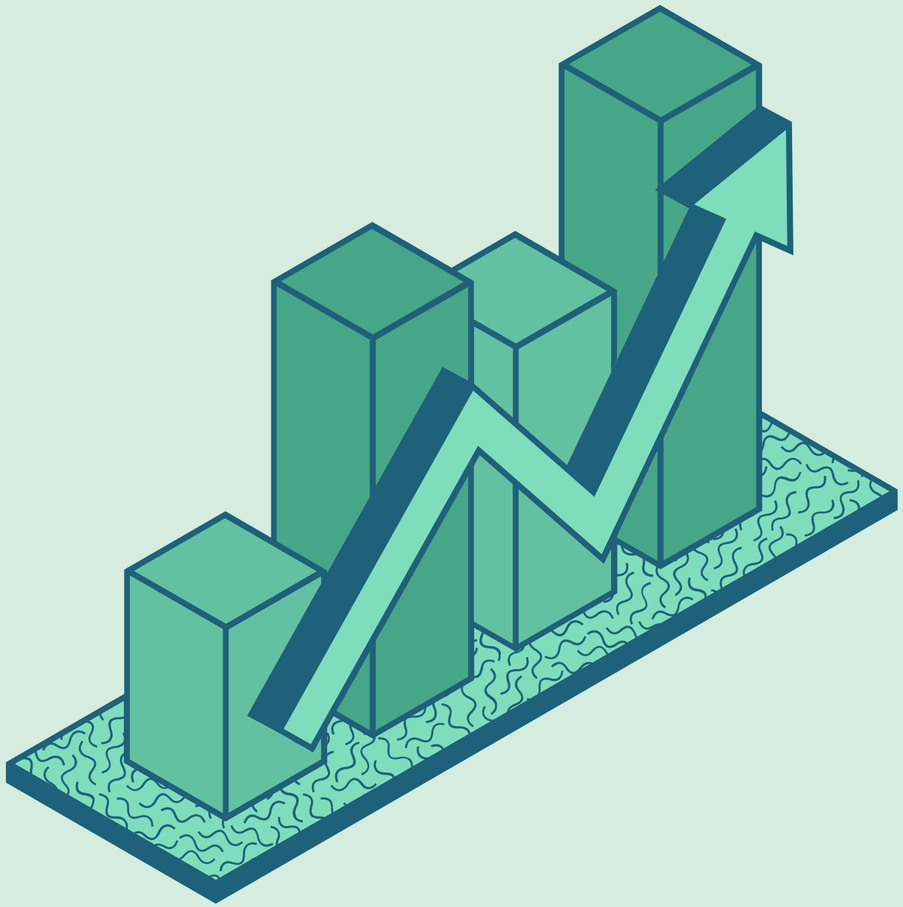


Identify important features in phishing classification

92% Accuracy Score

93% Recall Score

Proposals

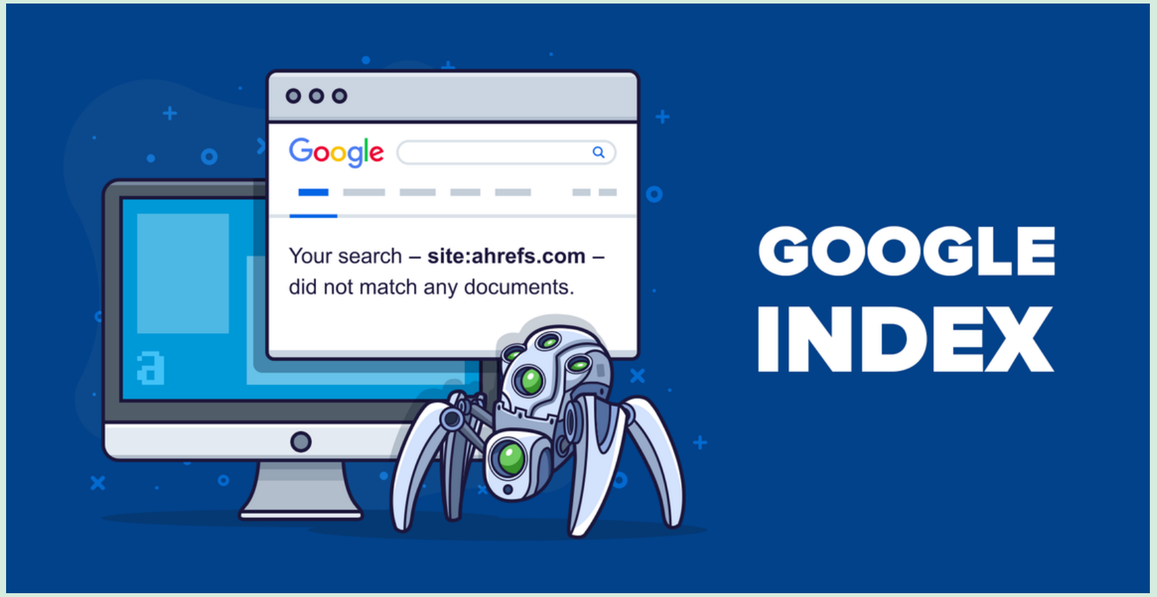


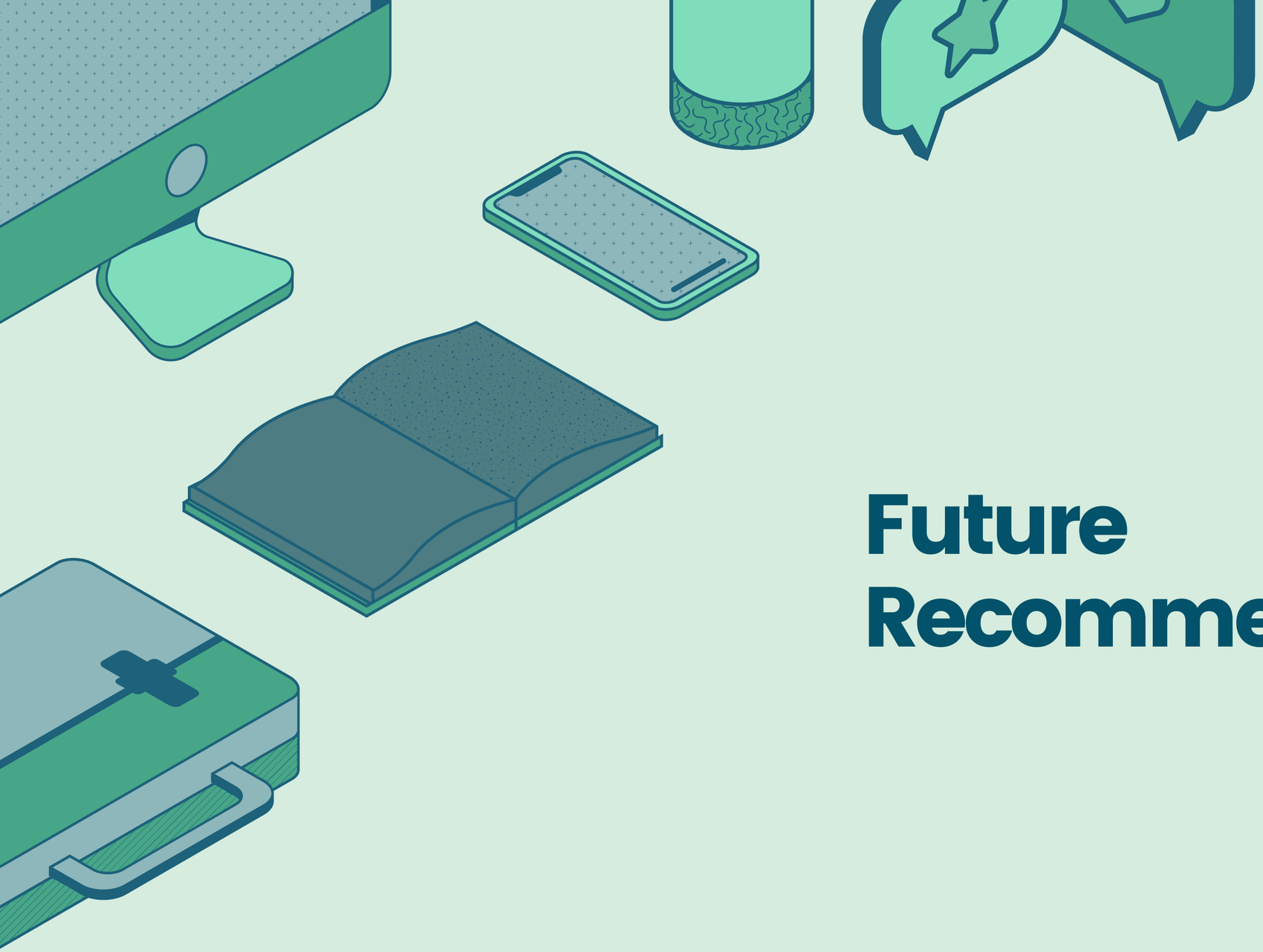
Proposals

Google Index accessibility

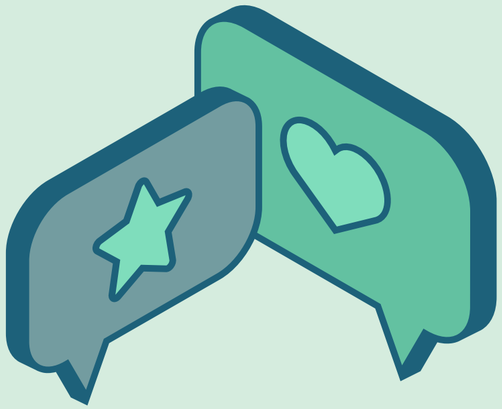
Educate on common phishing features

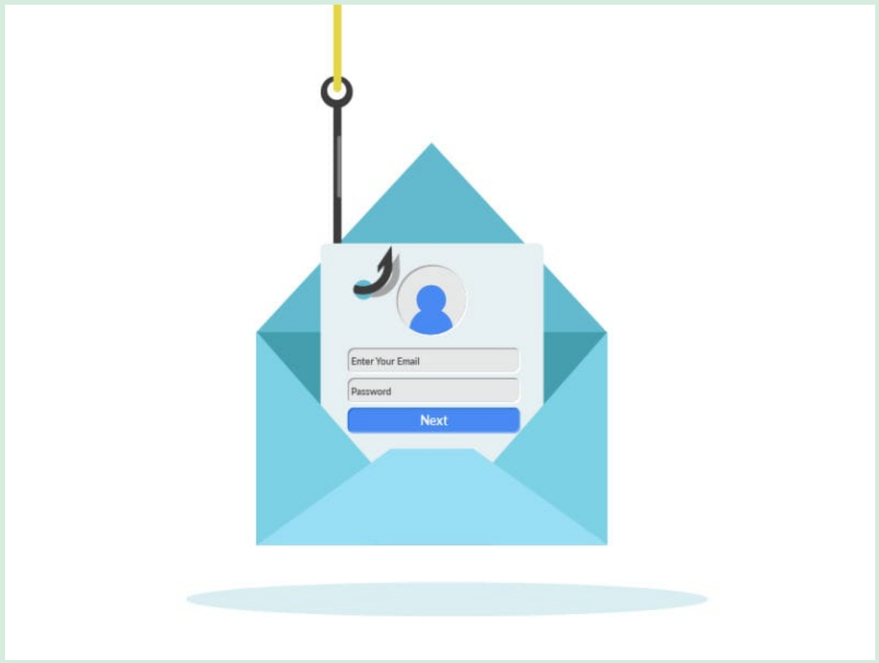
Stay up to date with new phishing techniques





Future Recommendations

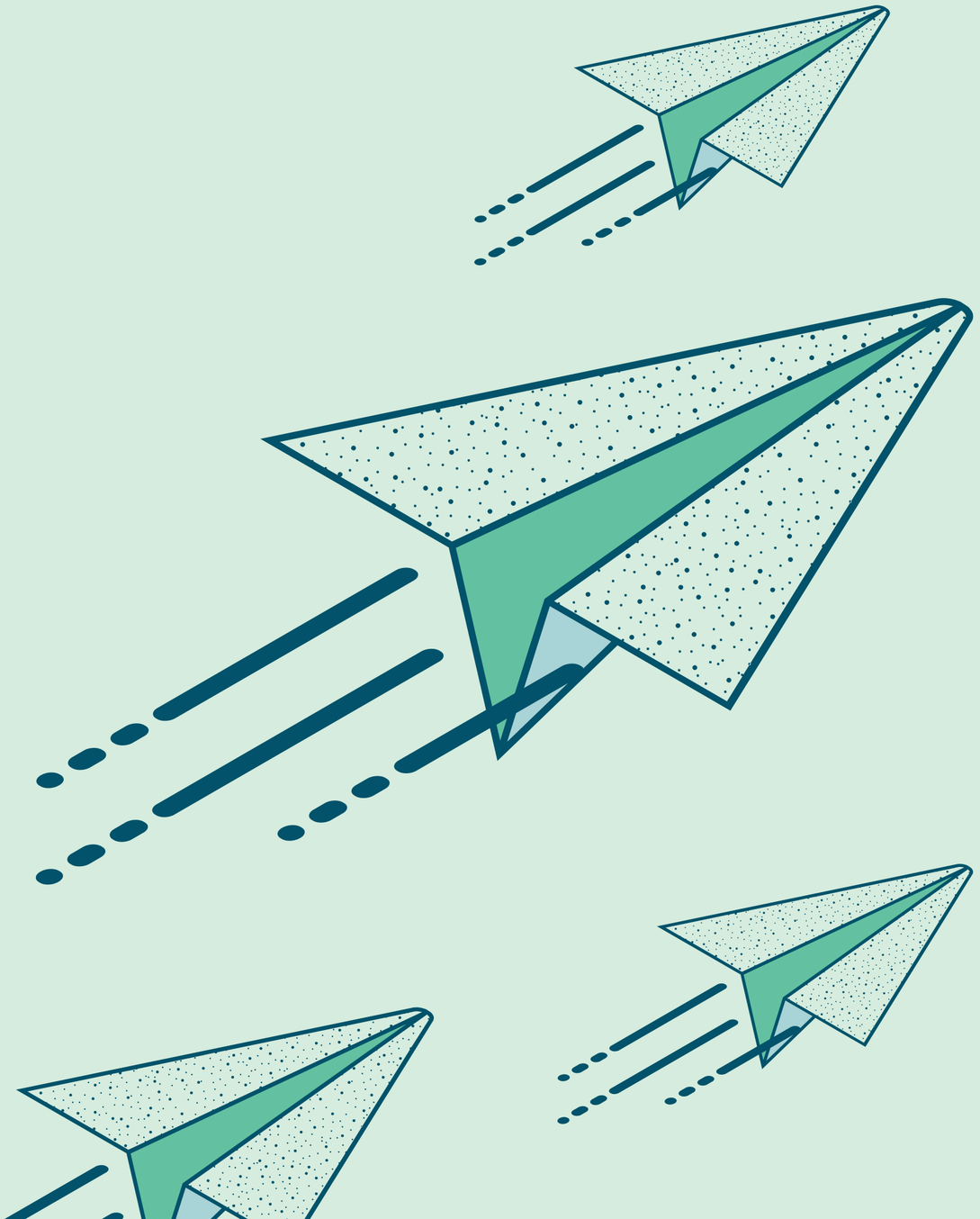




What's Next?

- Expand to text messaging and scam calls
- Create a website for phishing detection
- Utilize this model for other languages

Questions?



Thank you!

Peter Vuong
GitHub
LinkedIn