

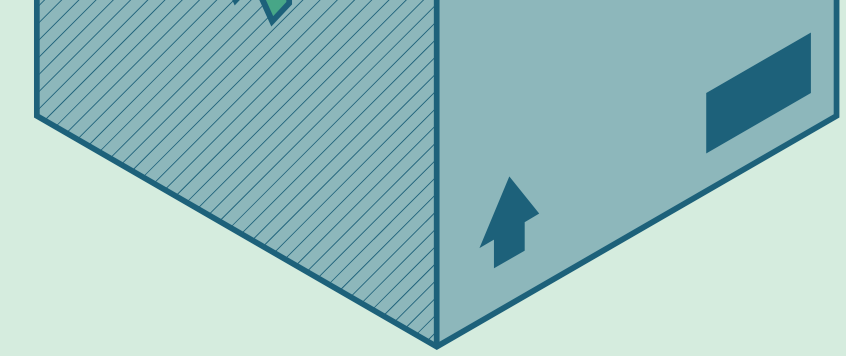
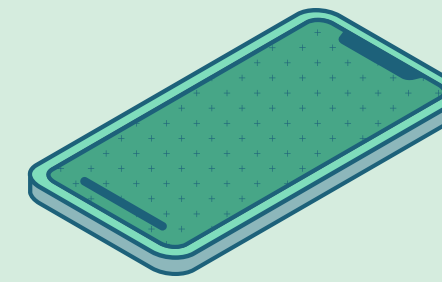
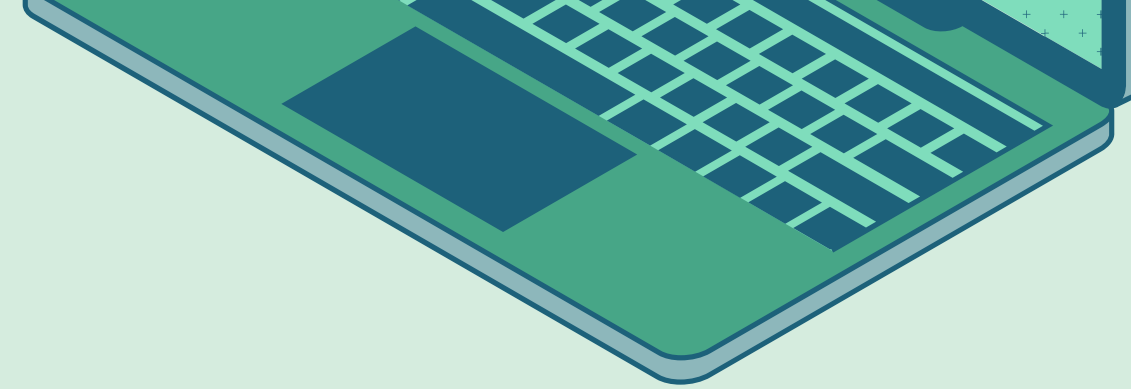
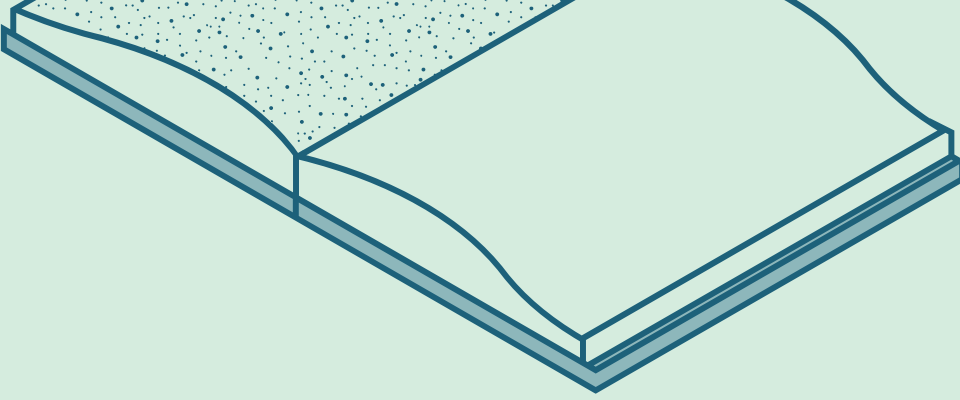


APWG

www.apwg.org

Web Phishing Detector

By Peter Vuong



Introduction



Peter Vuong

 [GitHub](#)

 [LinkedIn](#)

Agenda



Data



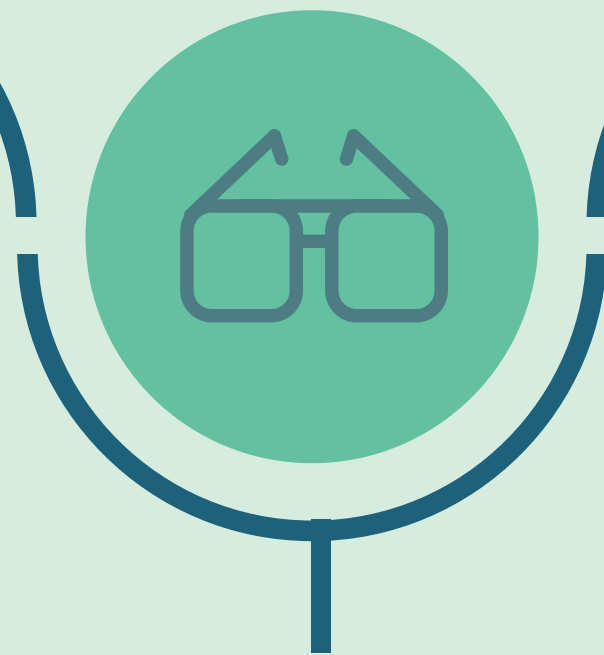
Proposals



Business
Understanding



Modeling &
Evaluation

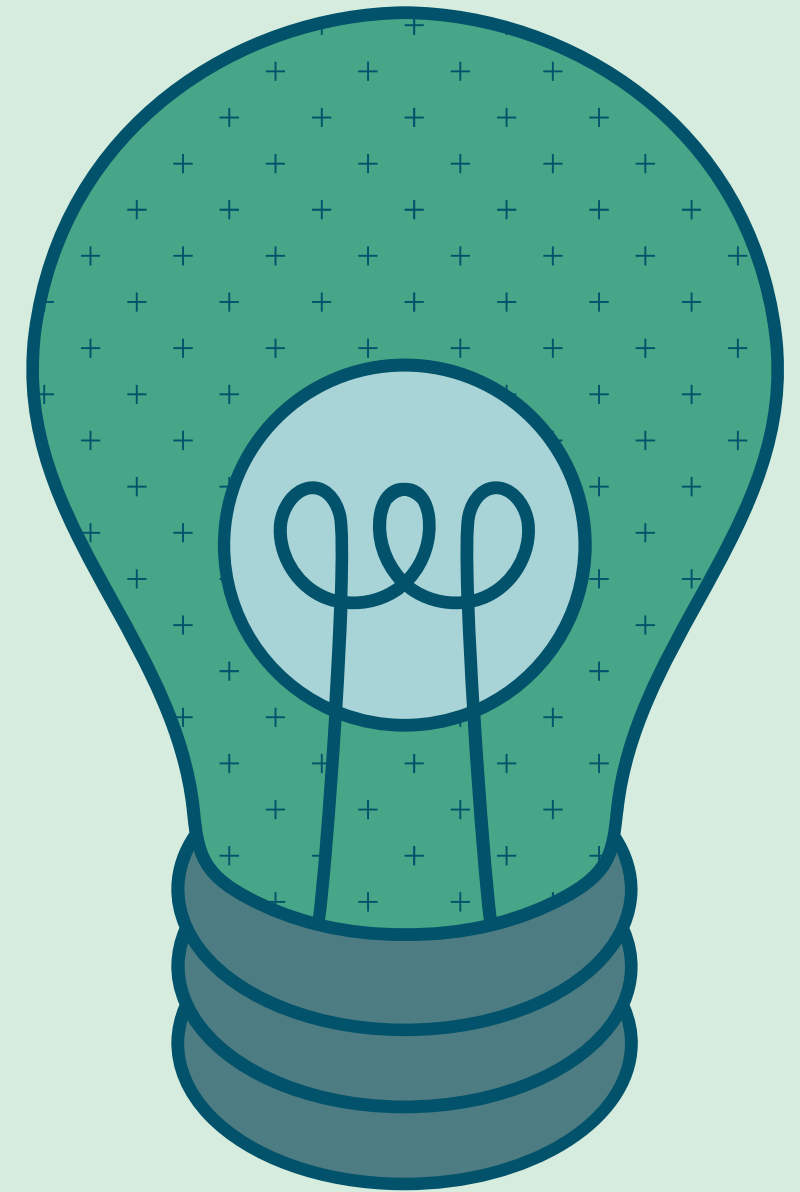


Future
Considerations



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

– APWG Mission Statement

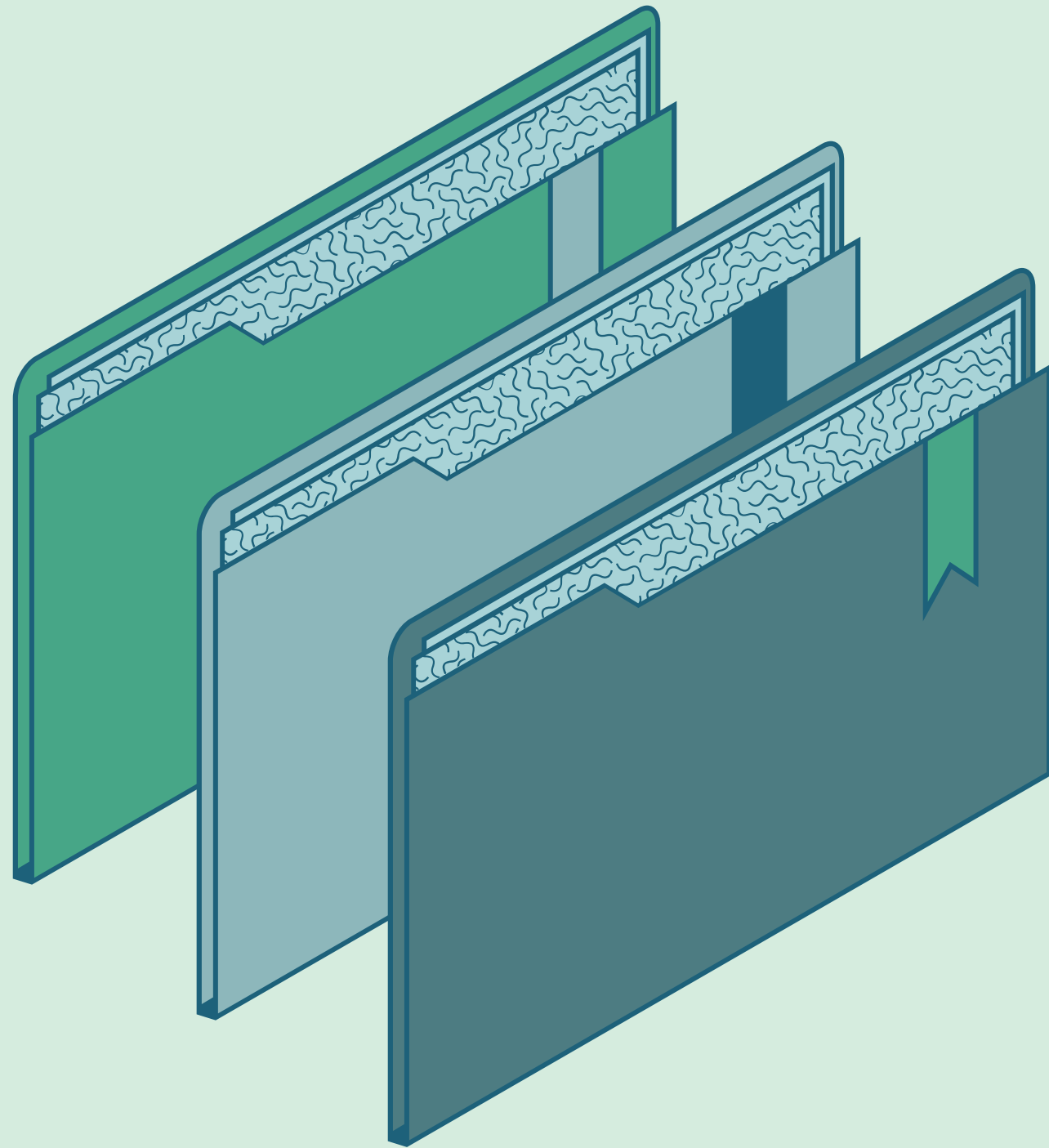




What is phishing?

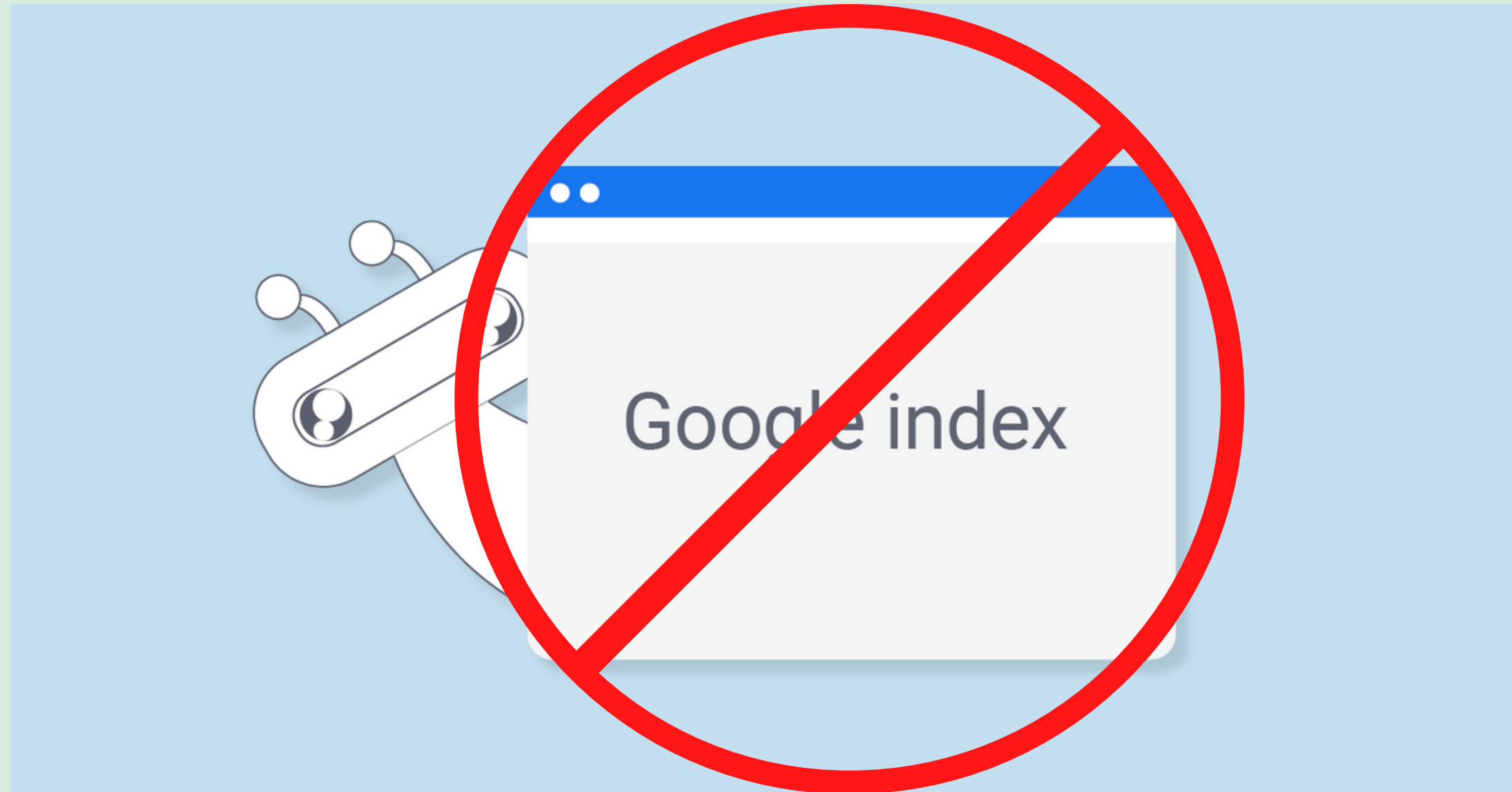
Socially engineered cyber attack

- Targets older people
- Sensitive information
- >200,000 victims in 2020

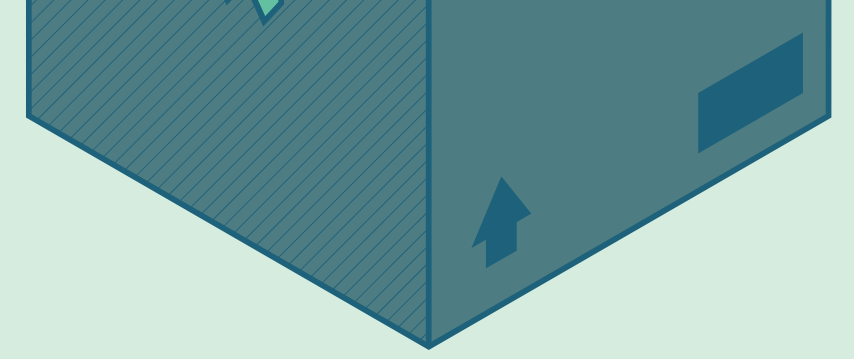
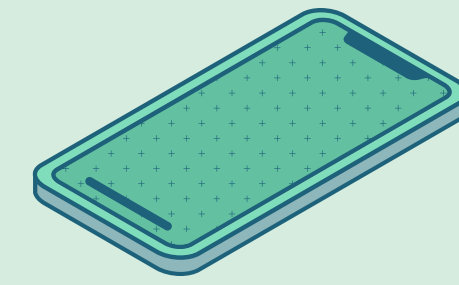
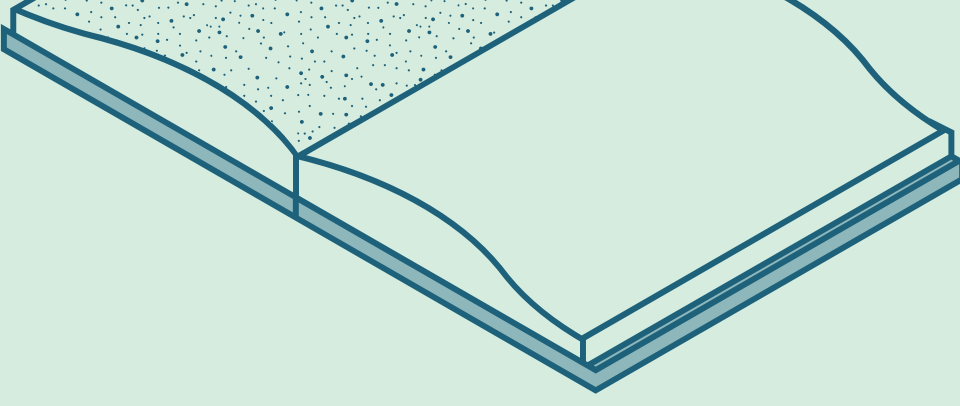


Bottom Line Up Front

- Accuracy = 92%
- Recall = 93%



20x more likely to be
phishing

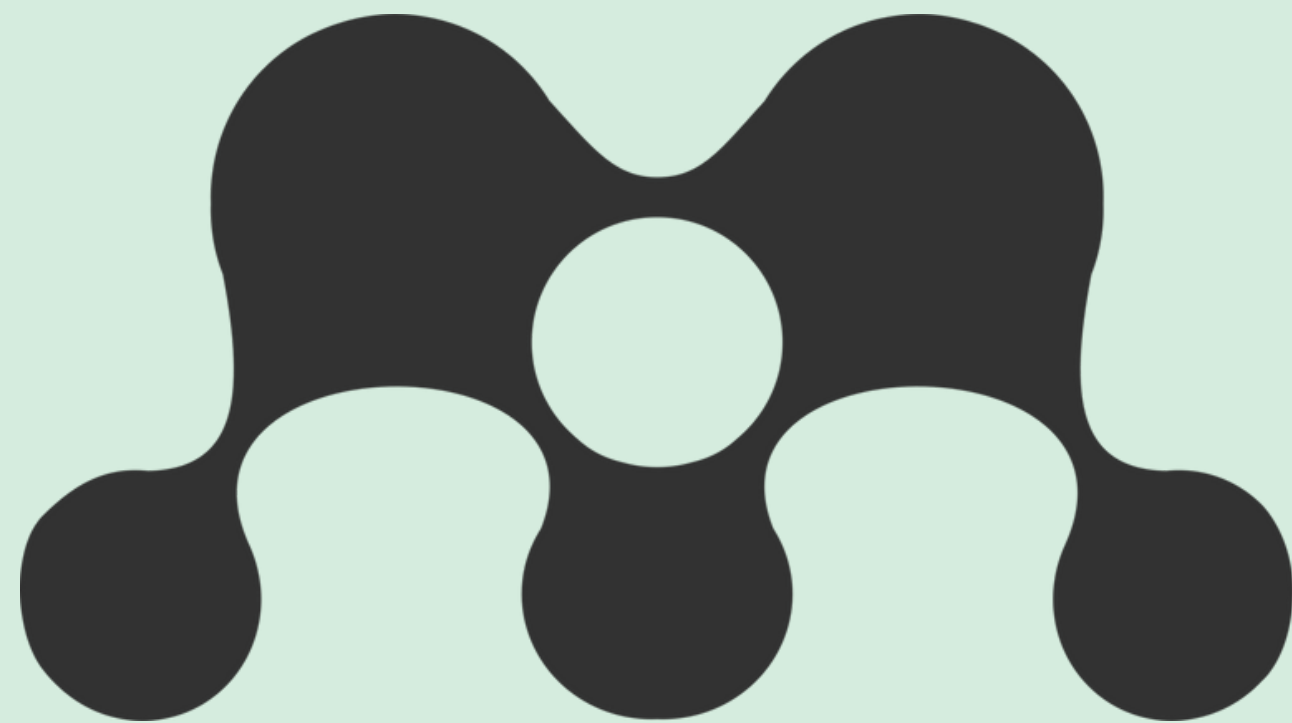


includes
signin content
admin plugins view
myaccount
alibaba dropbox site
themes images
login css

5x more likely to be
phishing



Data

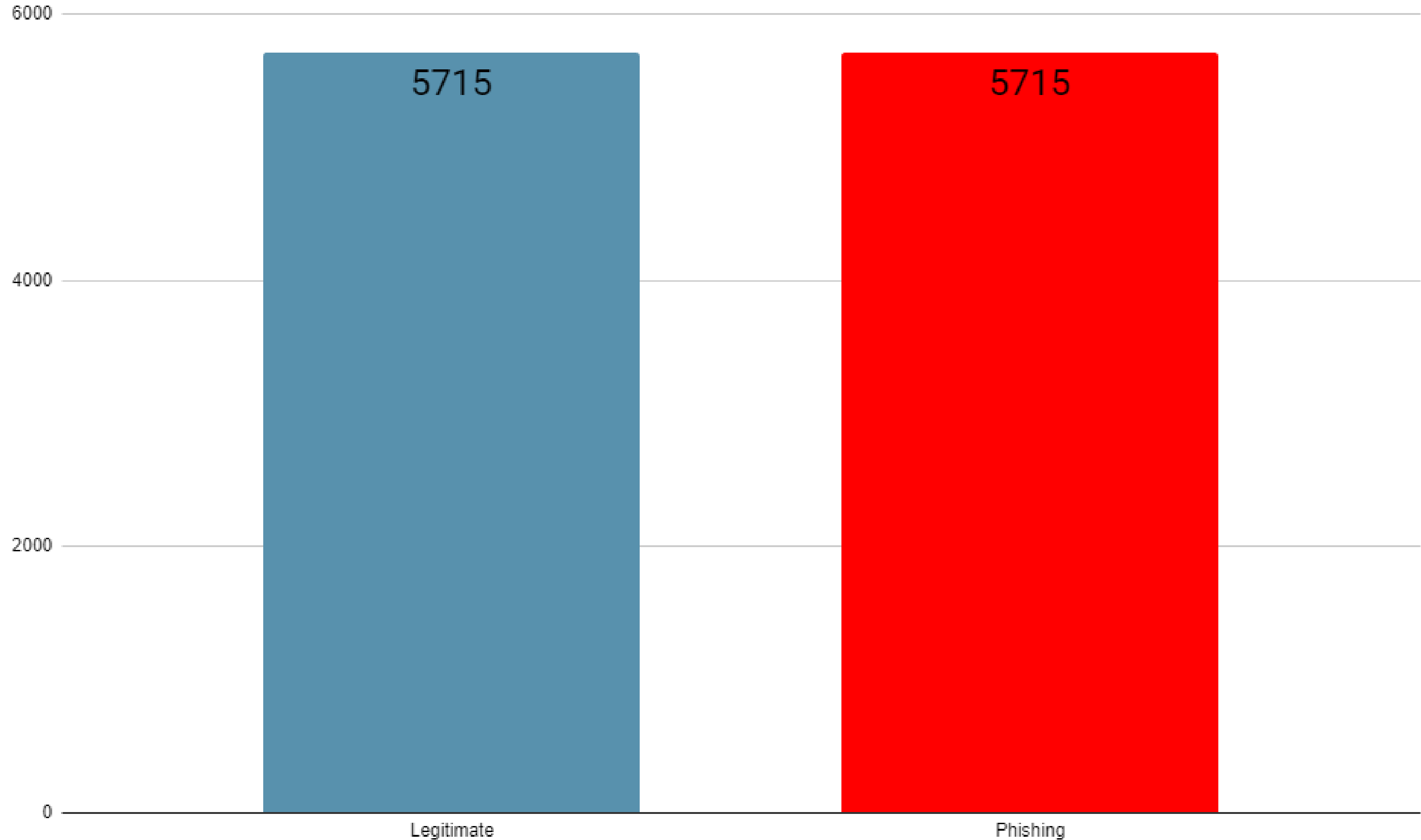


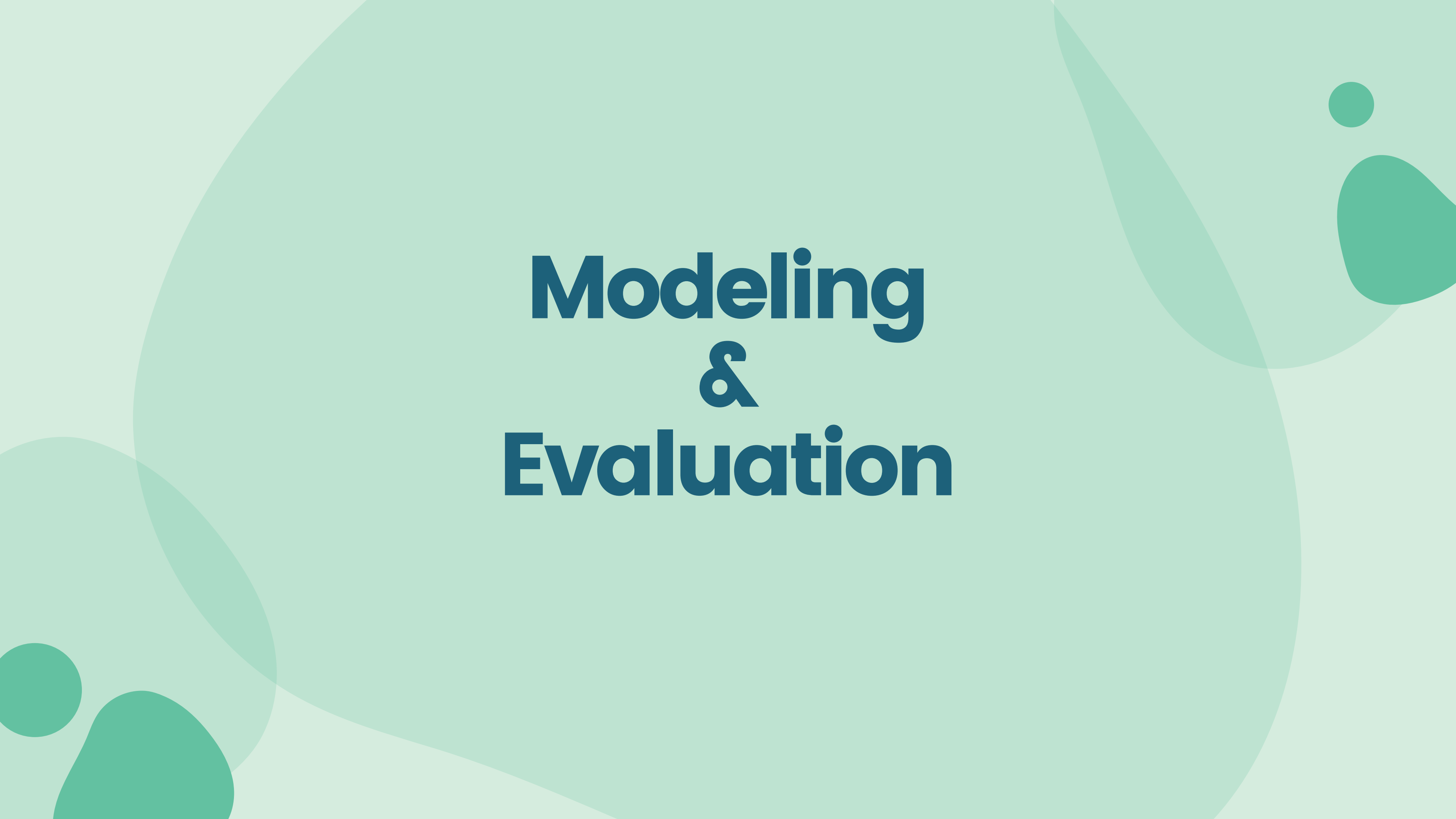
Mendeley Data

Published Jun 25, 2021

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

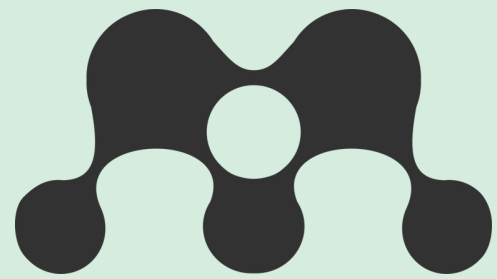
Distribution of Phishing vs. Legitimate Data





Modeling & Evaluation

Logistic Regression



Preprocessing



Logistic
Regression
with
ExtraTrees
Classifier



Phishing Detection

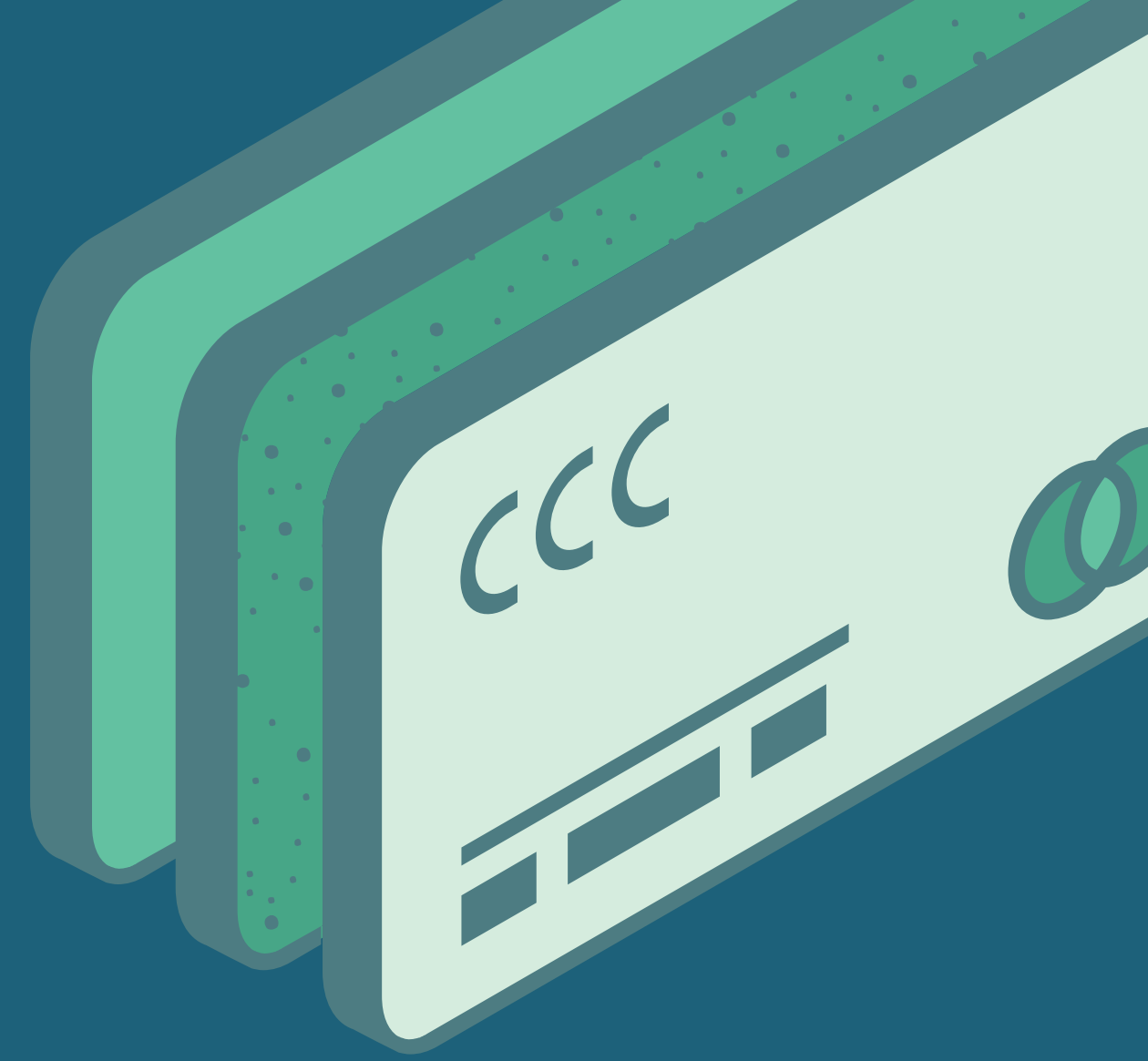


Logistic Regression

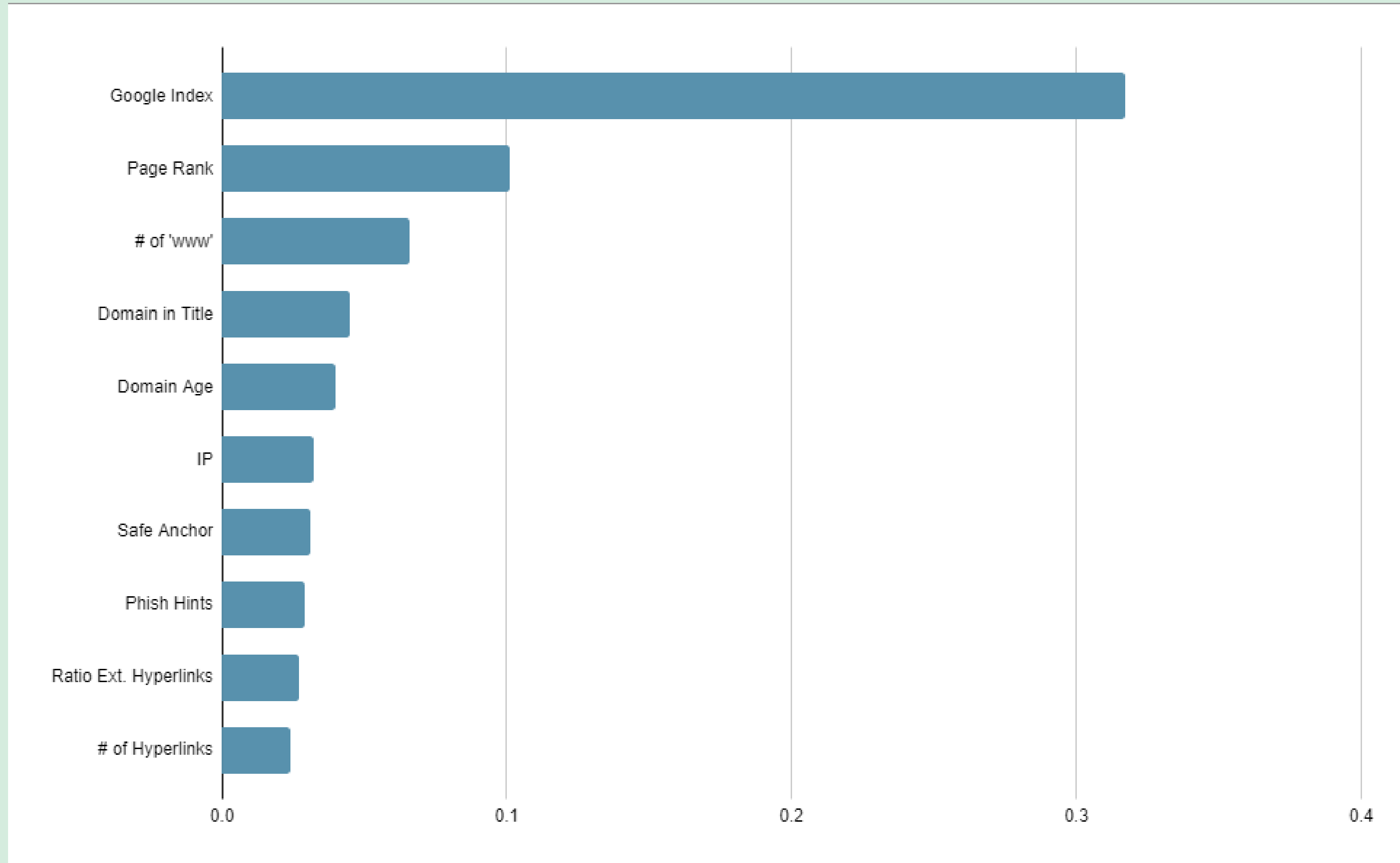
Focus on top 10 features

Accuracy Score = 92%

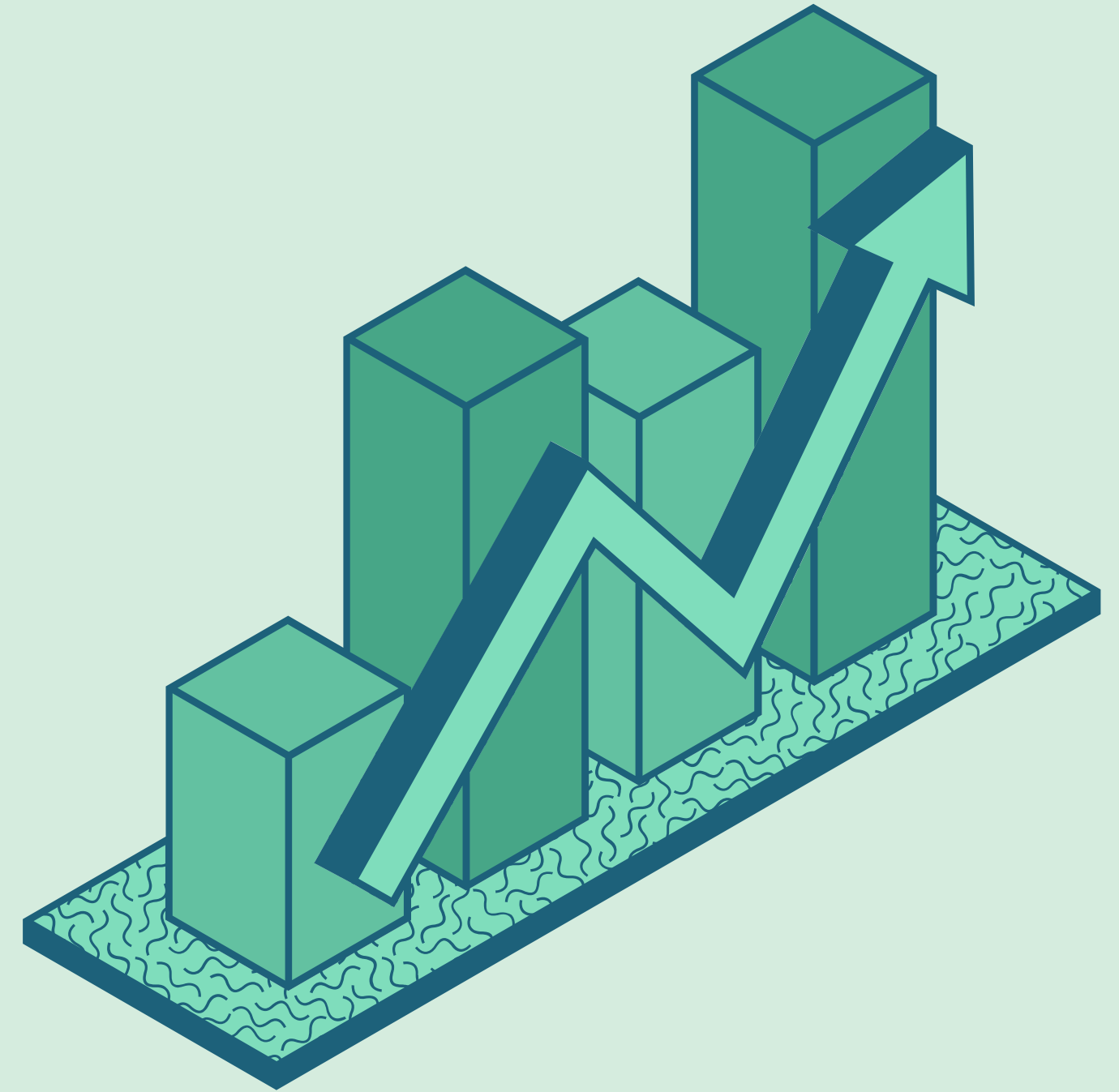
Recall Score = 93%



Top 10 Features from ExtraTreeClassifier



Proposals

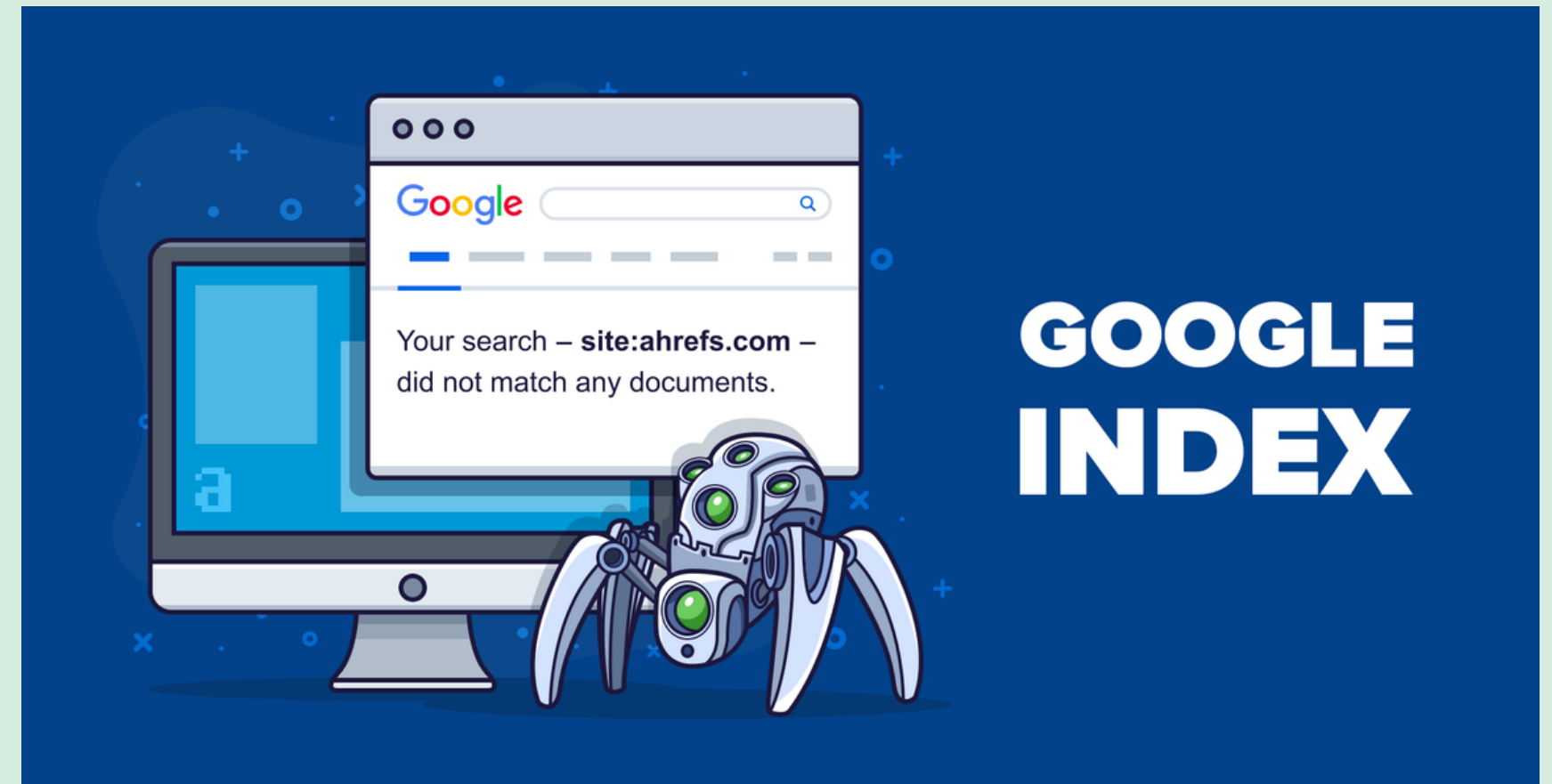


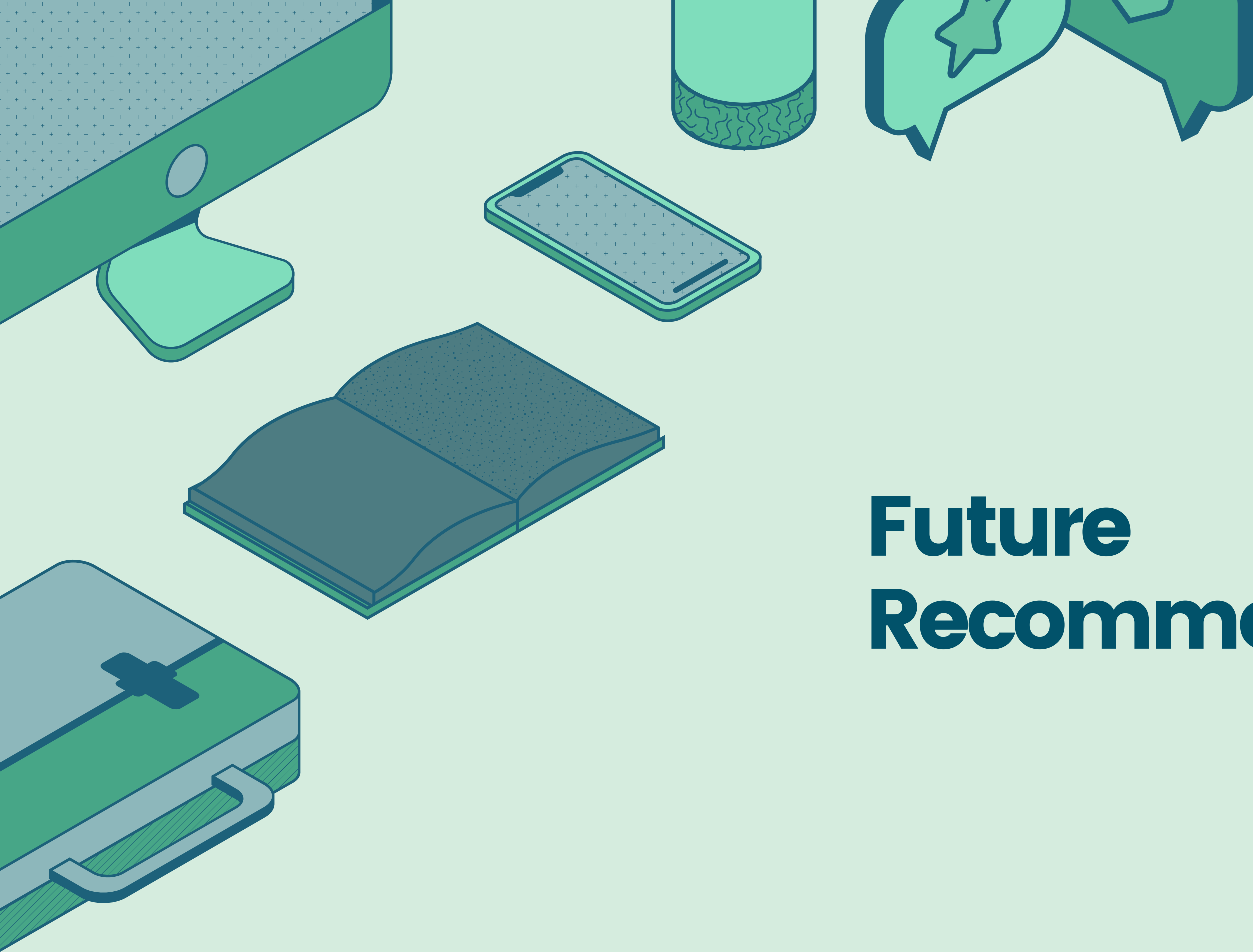
Proposals

Google Index accessibility

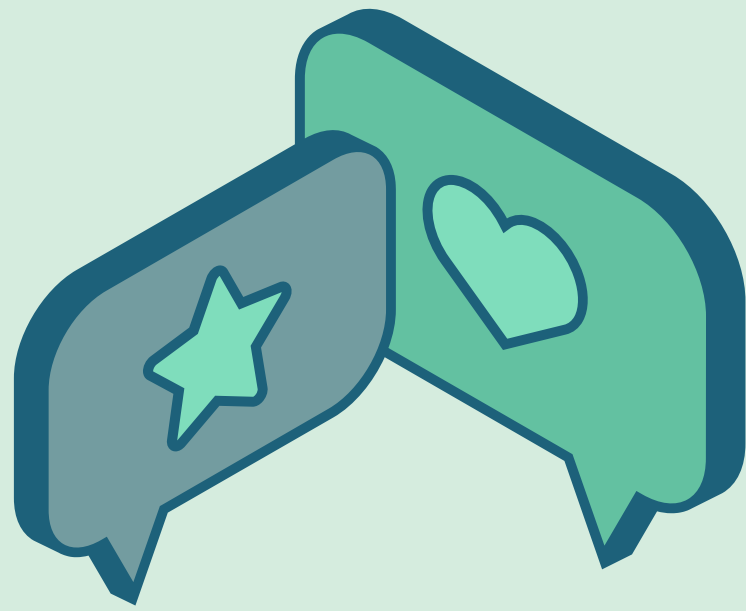
Educate on phishing features

Address new phishing methods



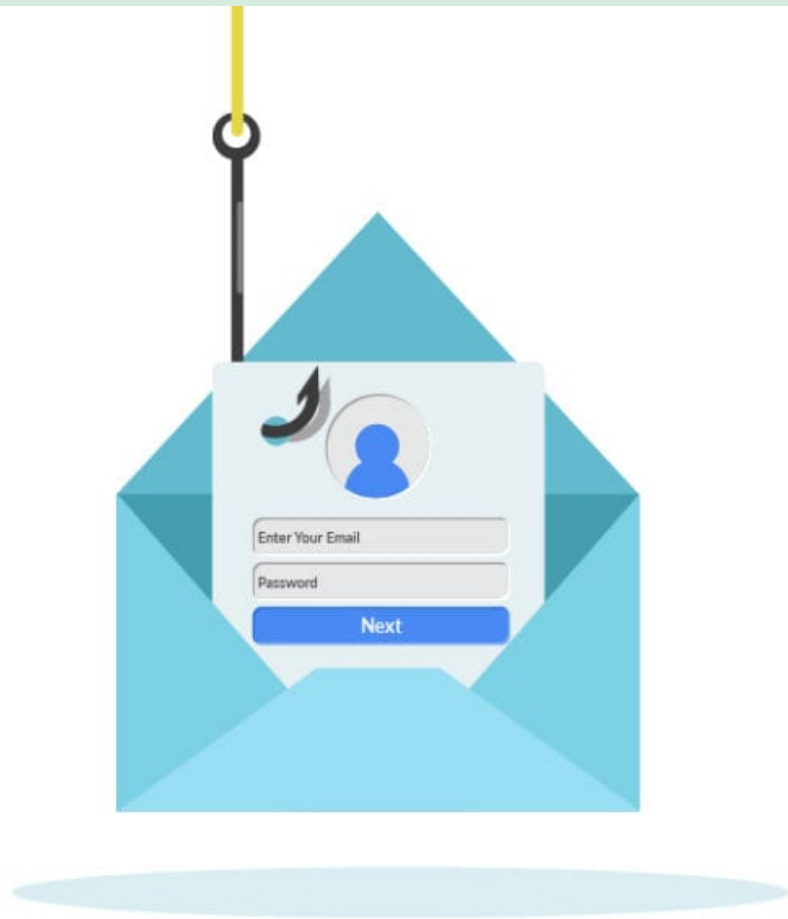


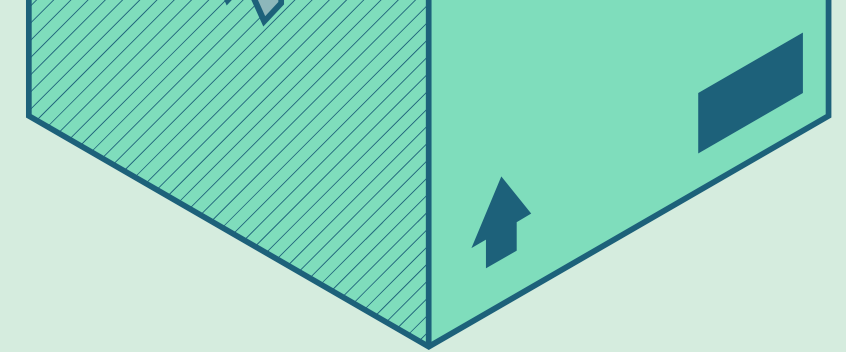
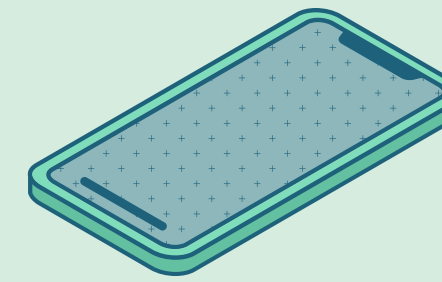
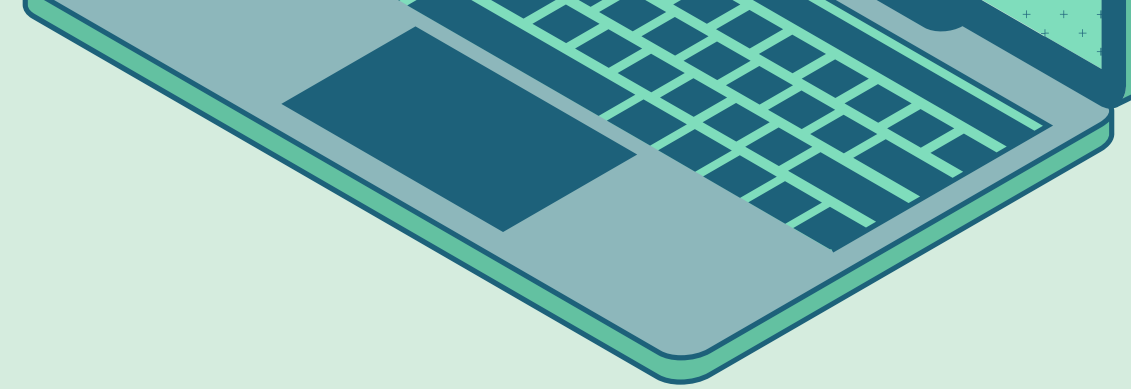
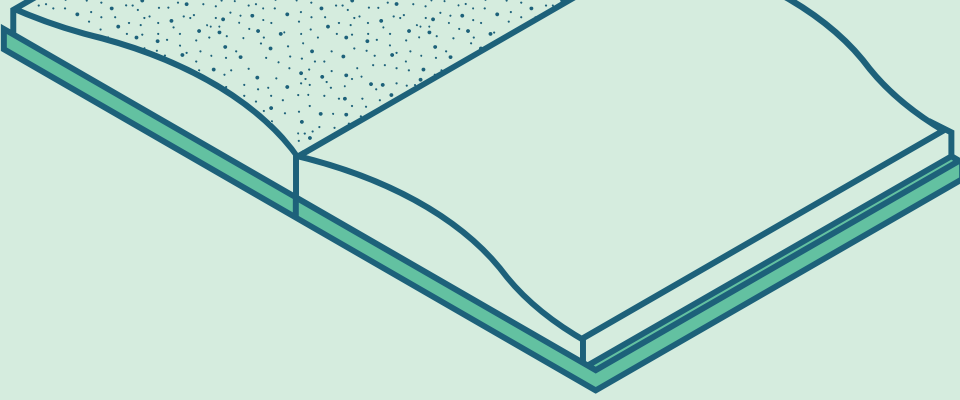
Future Recommendations



What's Next?

- Text messaging and scam calls
- Web program for phishing detection
- Model for other languages





Thank you!



Peter Vuong

 [GitHub](#)

 [LinkedIn](#)