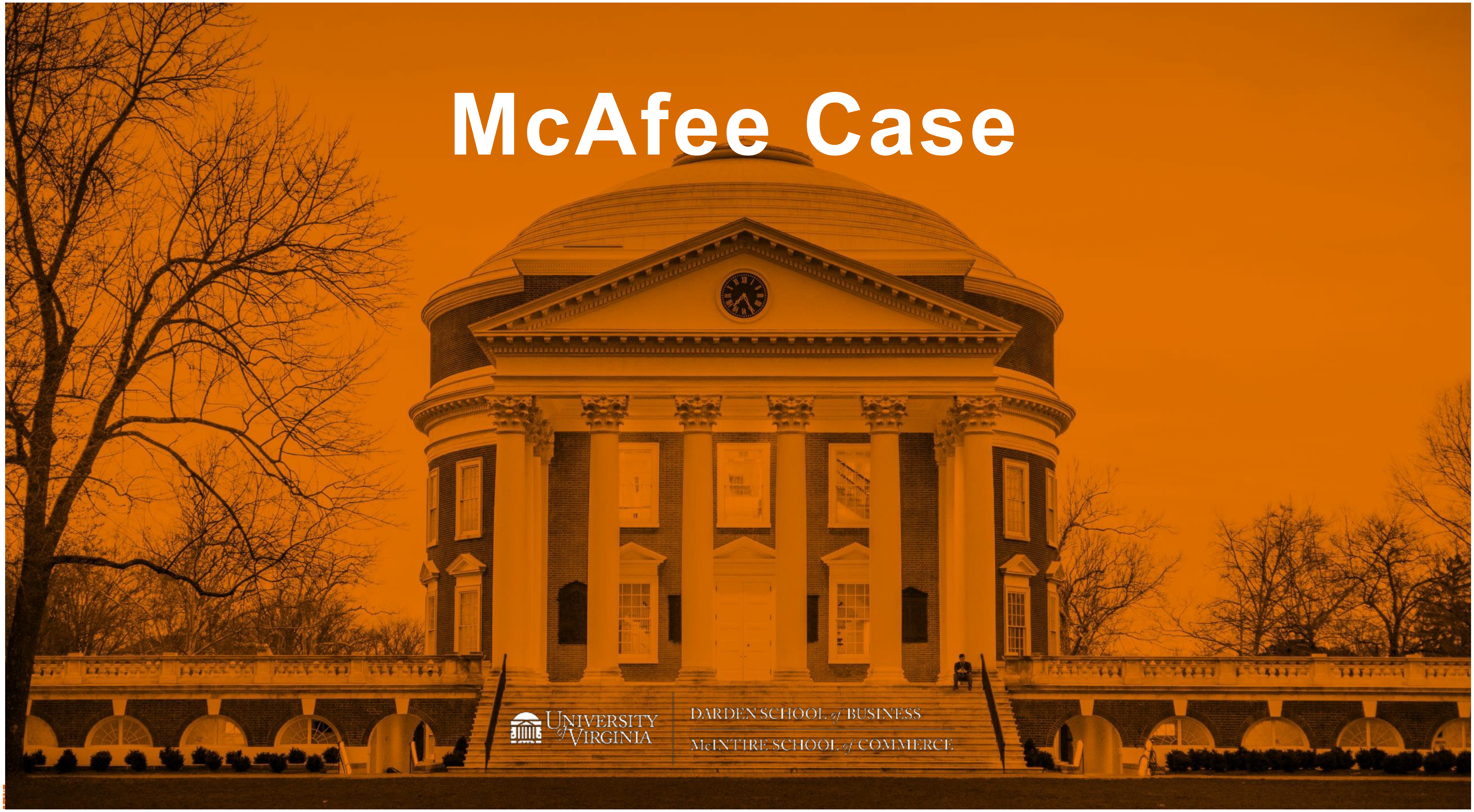
McAfee Case



UNIVERSITY of VIRGINIA

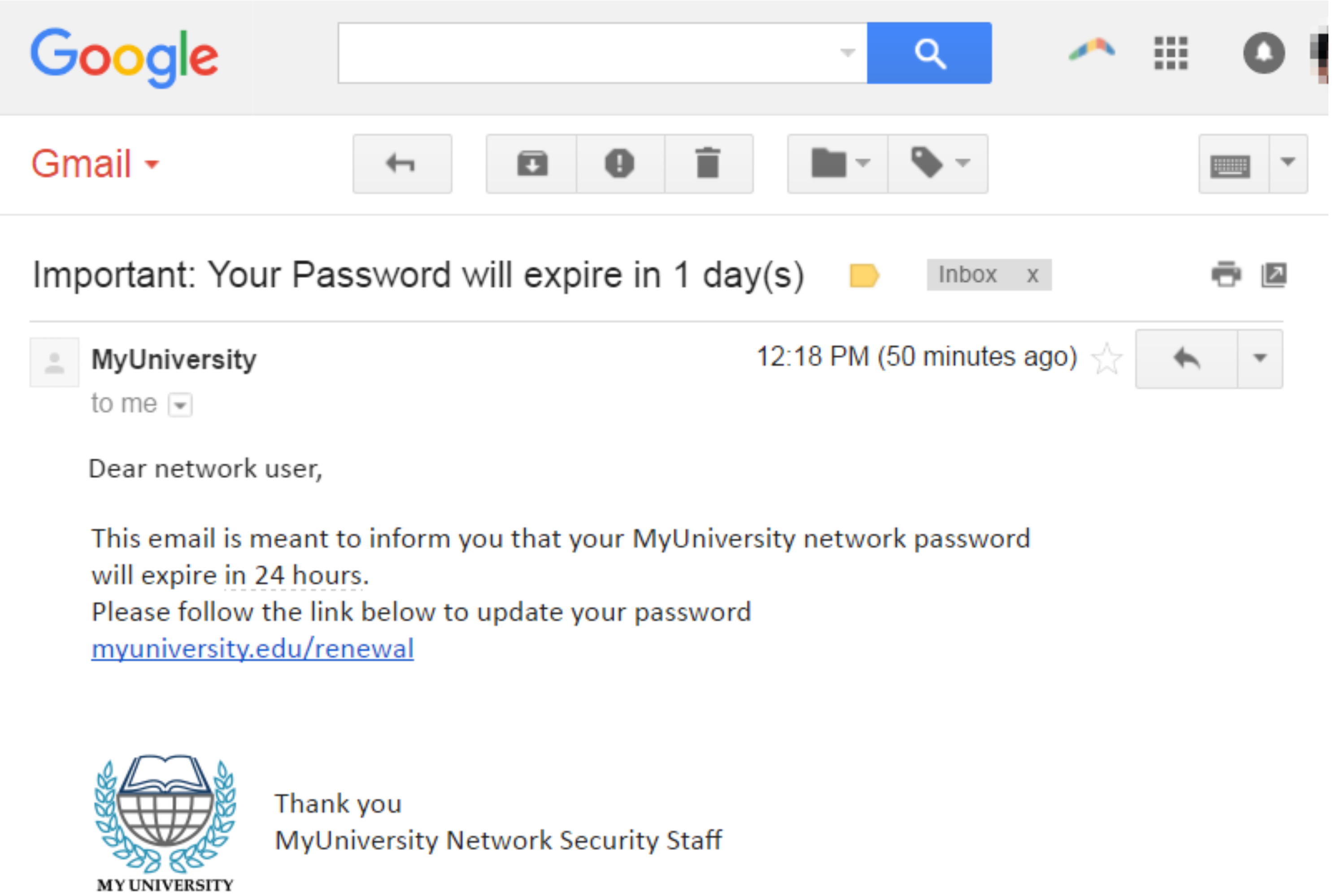
DARDEN SCHOOL of BUSINESS

McINTIRE SCHOOL of COMMERCE

MCAFEE CASE

AGENDA

- Case Background
  - Problem Statement
  - Why it matters, who cares?
  - Why should we build a great model?
- Notebook Review
- Case Competition
- Announce winners and wrap up



Important: Your Password will expire in 1 day(s)

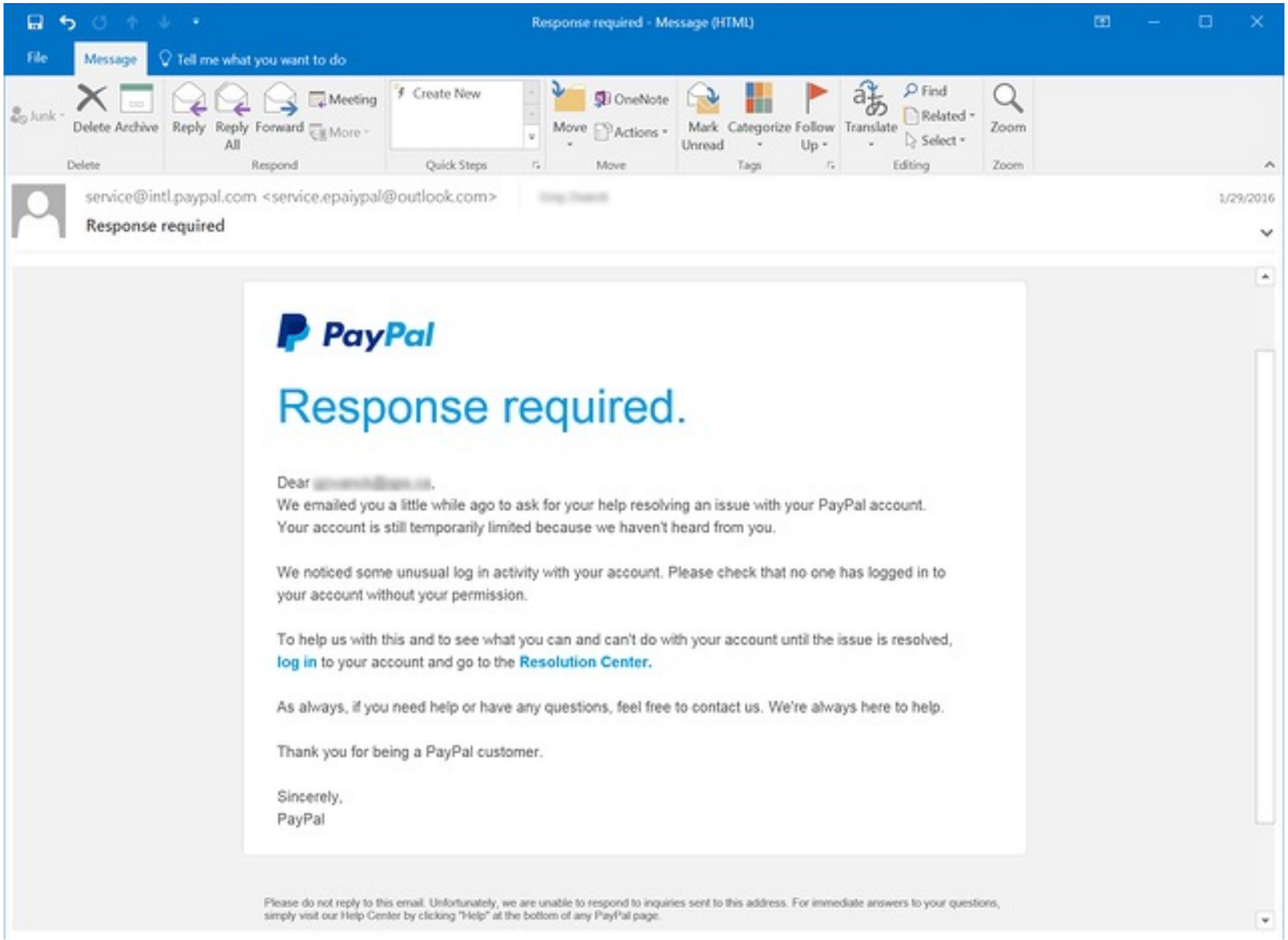
MyUniversity
to me

12:18 PM (50 minutes ago)

Dear network user,

This email is meant to inform you that your MyUniversity network password will expire in 24 hours.
Please follow the link below to update your password
myuniversity.edu/renewal

Thank you
MyUniversity Network Security Staff



service@intl.paypal.com <service.epaiypal@outlook.com> [redacted]

1/29/2016

Response required

PayPal

Response required.

Dear [redacted],
We emailed you a little while ago to ask for your help resolving an issue with your PayPal account. Your account is still temporarily limited because we haven't heard from you.

We noticed some unusual log in activity with your account. Please check that no one has logged in to your account without your permission.

To help us with this and to see what you can and can't do with your account until the issue is resolved, **log in** to your account and go to the **Resolution Center.**

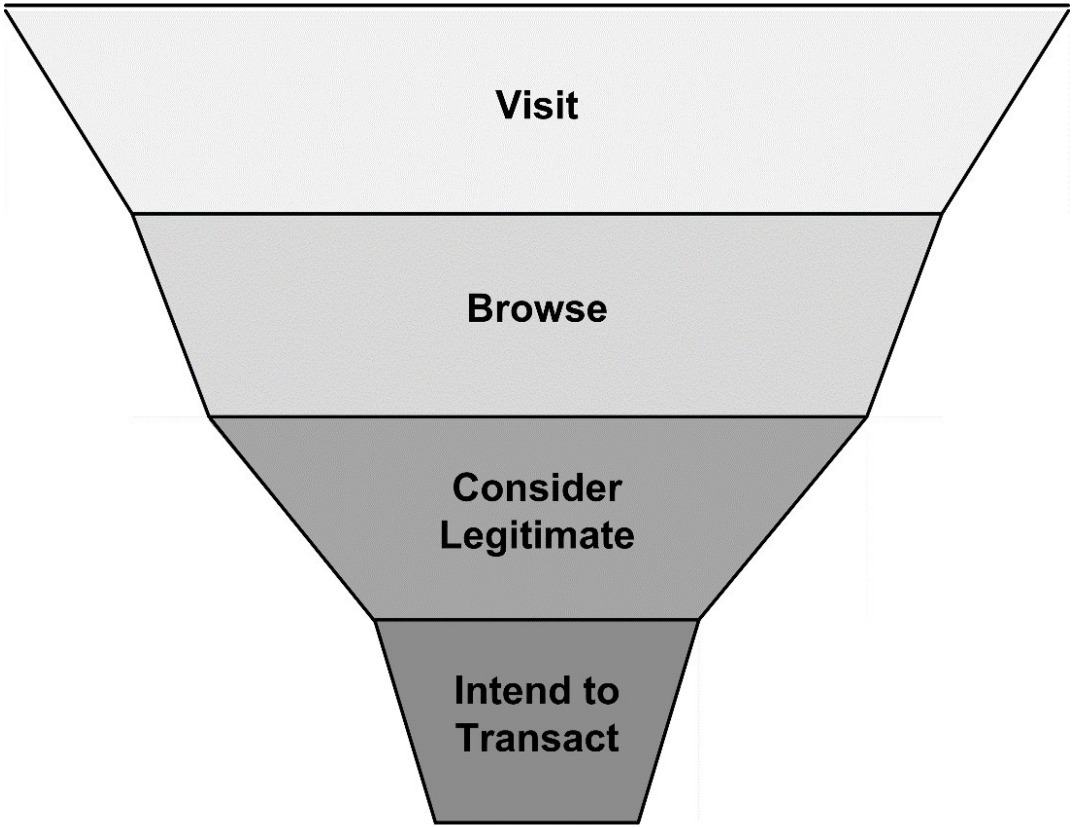
As always, if you need help or have any questions, feel free to contact us. We're always here to help.

Thank you for being a PayPal customer.

Sincerely,
PayPal

Please do not reply to this email. Unfortunately, we are unable to respond to inquiries sent to this address. For immediate answers to your questions, simply visit our Help Center by clicking "Help" at the bottom of any PayPal page.

Quick Background on the Case



Visit

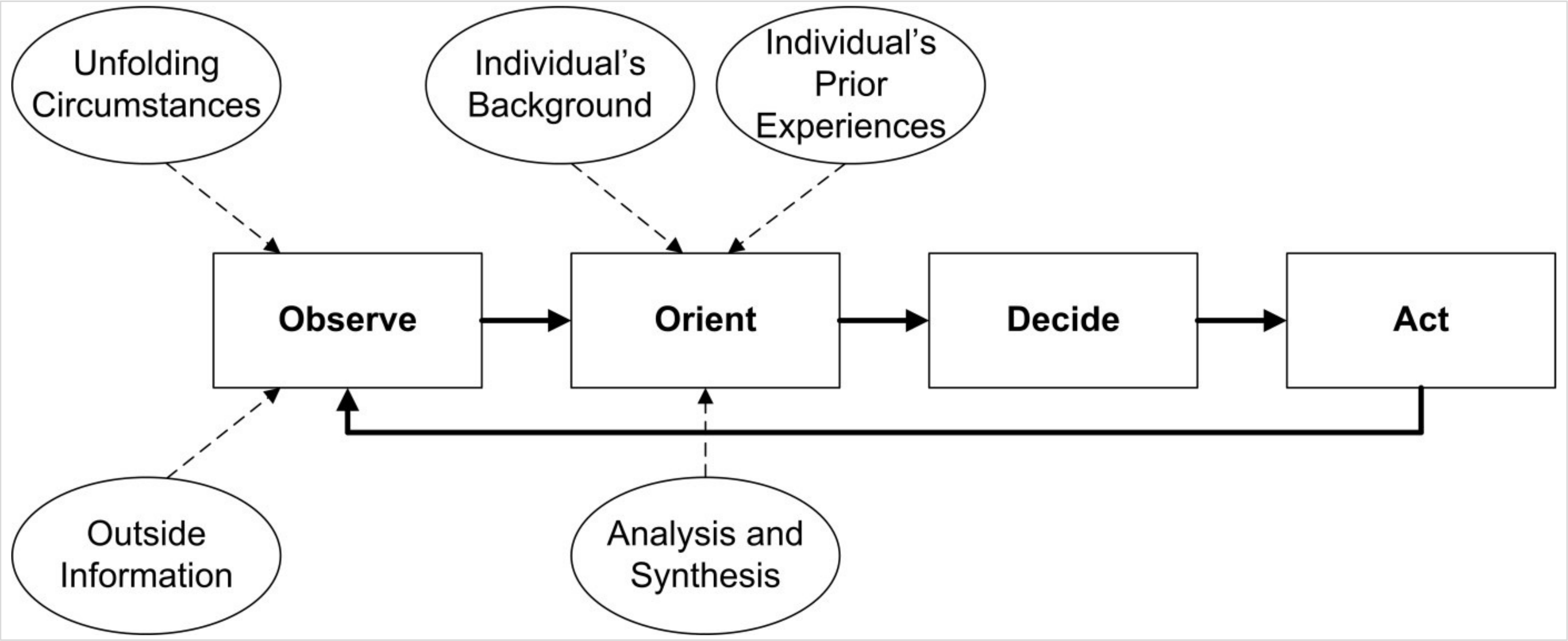
Browse

Consider Legitimate

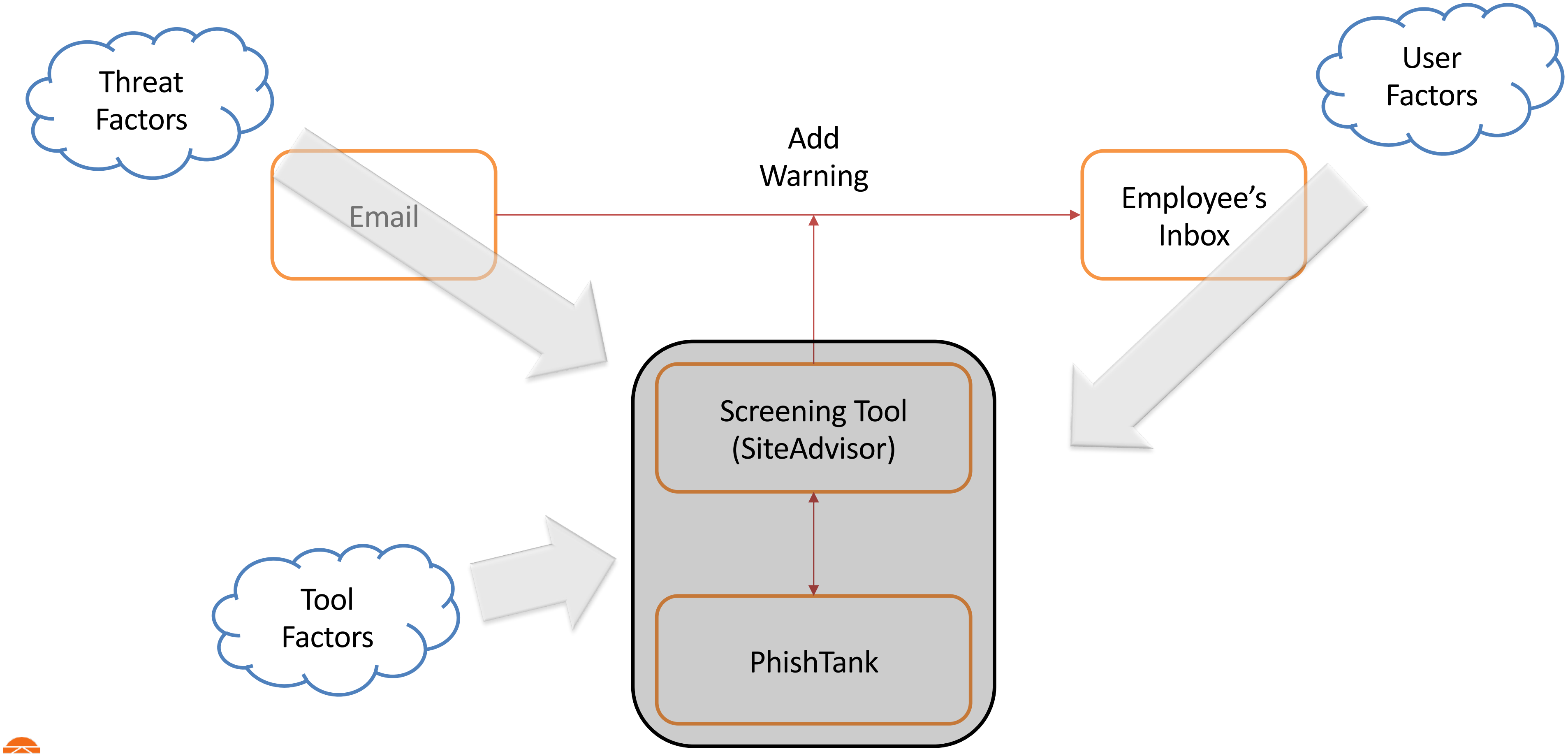
Intend to Transact

OODA: Observe, Orient, Decide, Act

- Understanding decision-making in risky, adversarial, real-time settings:



Status Quo


Threat Factors
Email
Add Warning
Employee’s Inbox
User Factors
Screening Tool (SiteAdvisor)
PhishTank
Tool Factors

Data & Funnel

Model Factor Categories and Funnel Stages

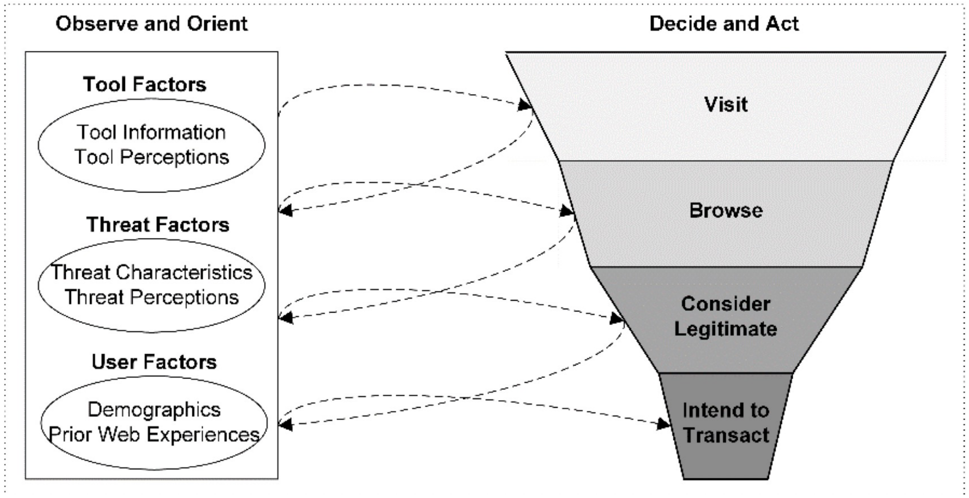
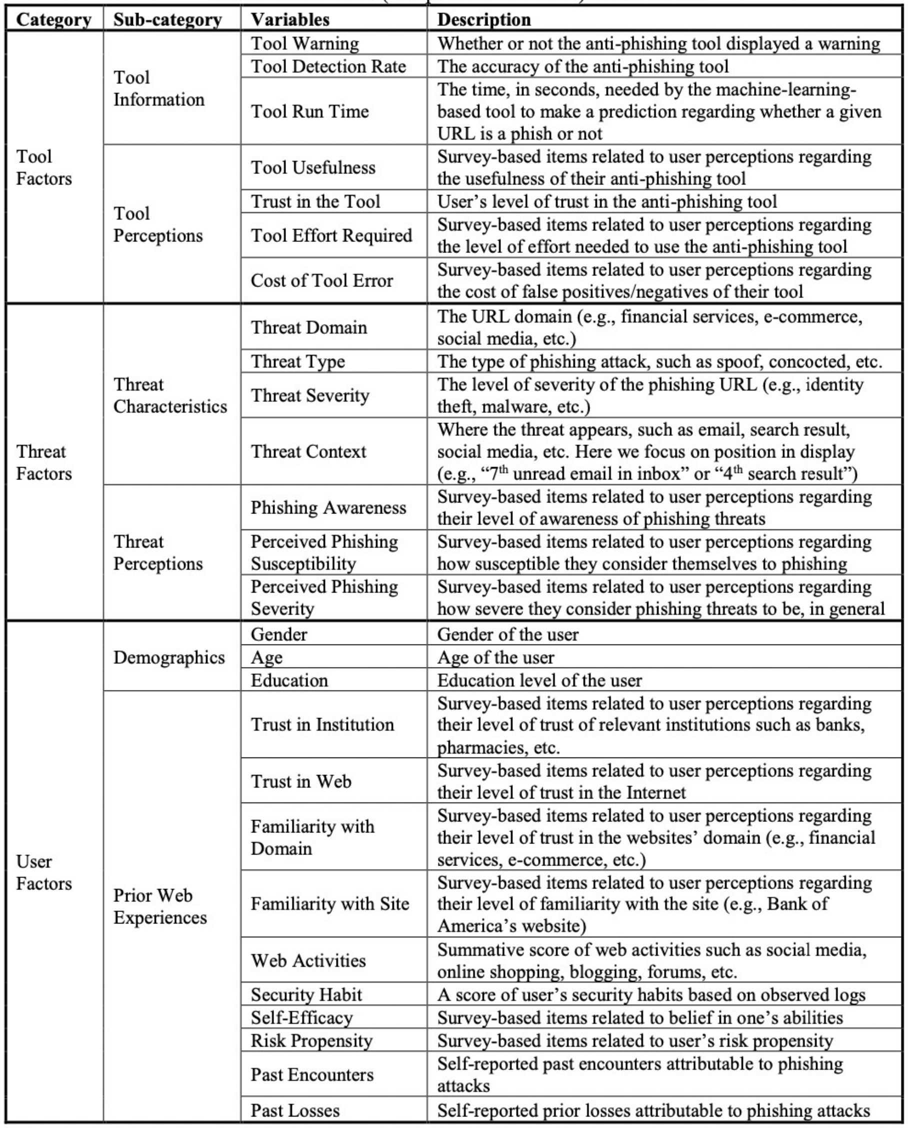
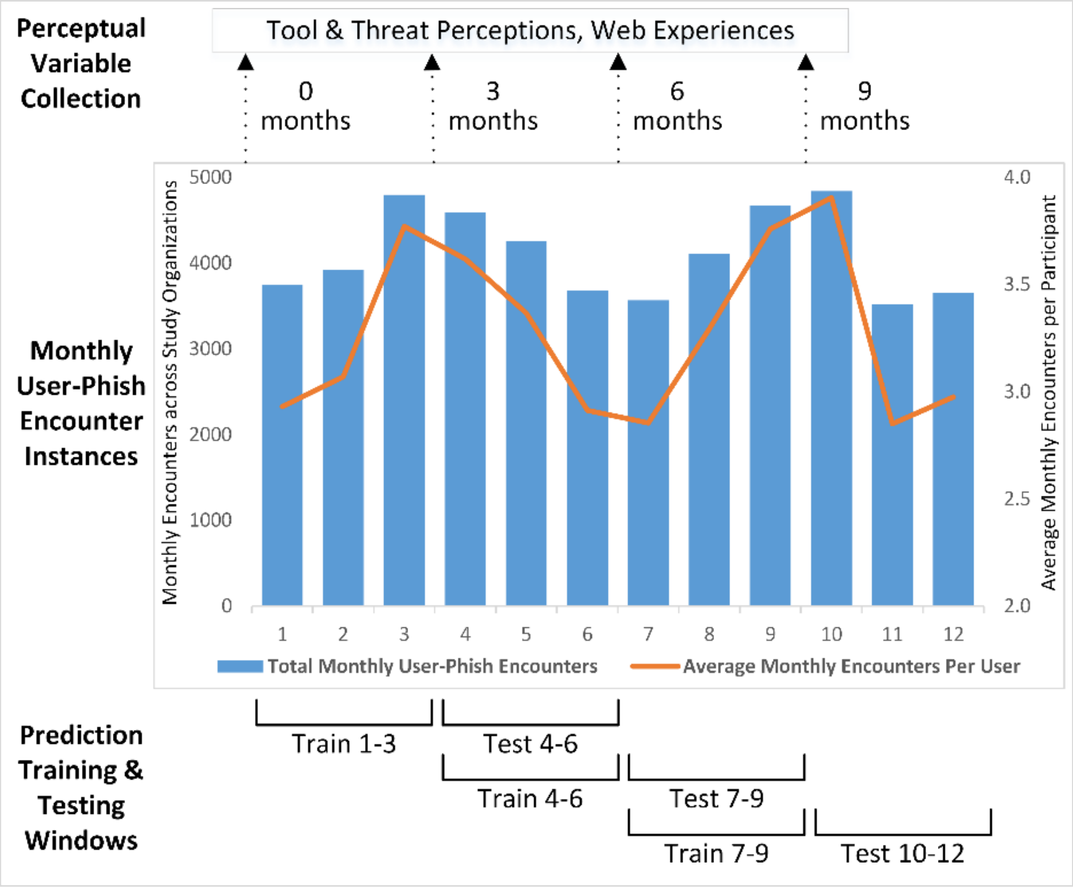


Table A1: Features (Independent Variables) used in PFM



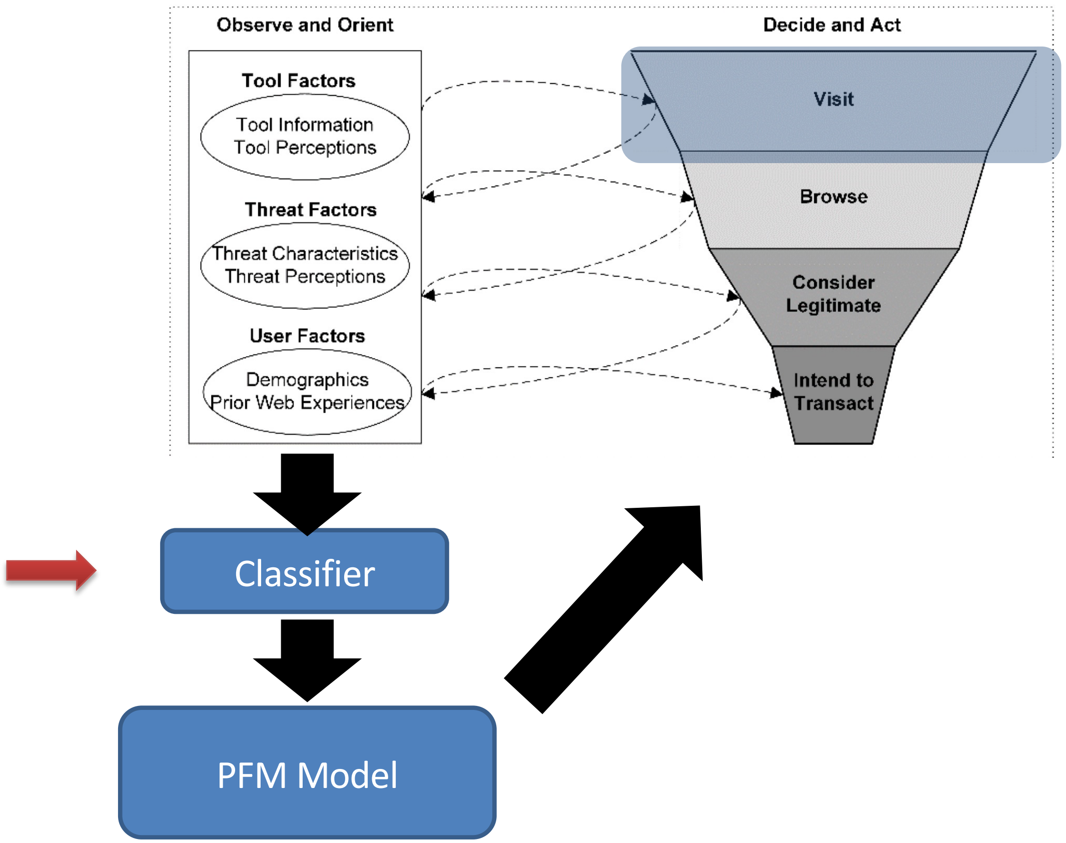
Category	Sub-category	Variables	Description
Tool Factors	Tool Information	Tool Warning	Whether or not the anti-phishing tool displayed a warning
		Tool Detection Rate	The accuracy of the anti-phishing tool
		Tool Run Time	The time, in seconds, needed by the machine-learning-based tool to make a prediction regarding whether a given URL is a phish or not
	Tool Perceptions	Tool Usefulness	Survey-based items related to user perceptions regarding the usefulness of their anti-phishing tool
		Trust in the Tool	User's level of trust in the anti-phishing tool
		Tool Effort Required	Survey-based items related to user perceptions regarding the level of effort needed to use the anti-phishing tool
		Cost of Tool Error	Survey-based items related to user perceptions regarding the cost of false positives/negatives of their tool
Threat Factors	Threat Characteristics	Threat Domain	The URL domain (e.g., financial services, e-commerce, social media, etc.)
		Threat Type	The type of phishing attack, such as spoof, concocted, etc.
		Threat Severity	The level of severity of the phishing URL (e.g., identity theft, malware, etc.)
		Threat Context	Where the threat appears, such as email, search result, social media, etc. Here we focus on position in display (e.g., "7th unread email in inbox" or "4th search result")
	Threat Perceptions	Phishing Awareness	Survey-based items related to user perceptions regarding their level of awareness of phishing threats
		Perceived Phishing Susceptibility	Survey-based items related to user perceptions regarding how susceptible they consider themselves to phishing
		Perceived Phishing Severity	Survey-based items related to user perceptions regarding how severe they consider phishing threats to be, in general
User Factors	Demographics	Gender	Gender of the user
		Age	Age of the user
		Education	Education level of the user
	Prior Web Experiences	Trust in Institution	Survey-based items related to user perceptions regarding their level of trust of relevant institutions such as banks, pharmacies, etc.
		Trust in Web	Survey-based items related to user perceptions regarding their level of trust in the Internet
		Familiarity with Domain	Survey-based items related to user perceptions regarding their level of trust in the websites' domain (e.g., financial services, e-commerce, etc.)
		Familiarity with Site	Survey-based items related to user perceptions regarding their level of familiarity with the site (e.g., Bank of America's website)
		Web Activities	Summative score of web activities such as social media, online shopping, blogging, forums, etc.
		Security Habit	A score of user's security habits based on observed logs
		Self-Efficacy	Survey-based items related to belief in one's abilities
		Risk Propensity	Survey-based items related to user's risk propensity
		Past Encounters	Self-reported past encounters attributable to phishing attacks
		Past Losses	Self-reported prior losses attributable to phishing attacks

The Phishing Funnel Model: Field Experiment Setup (FinOrg)



The Phishing Funnel Model

Model Factor Categories and Funnel Stages



Classifier

PFM Model

Why the predictive performance of the model is important?

Case Competition

- Login to Canvas.

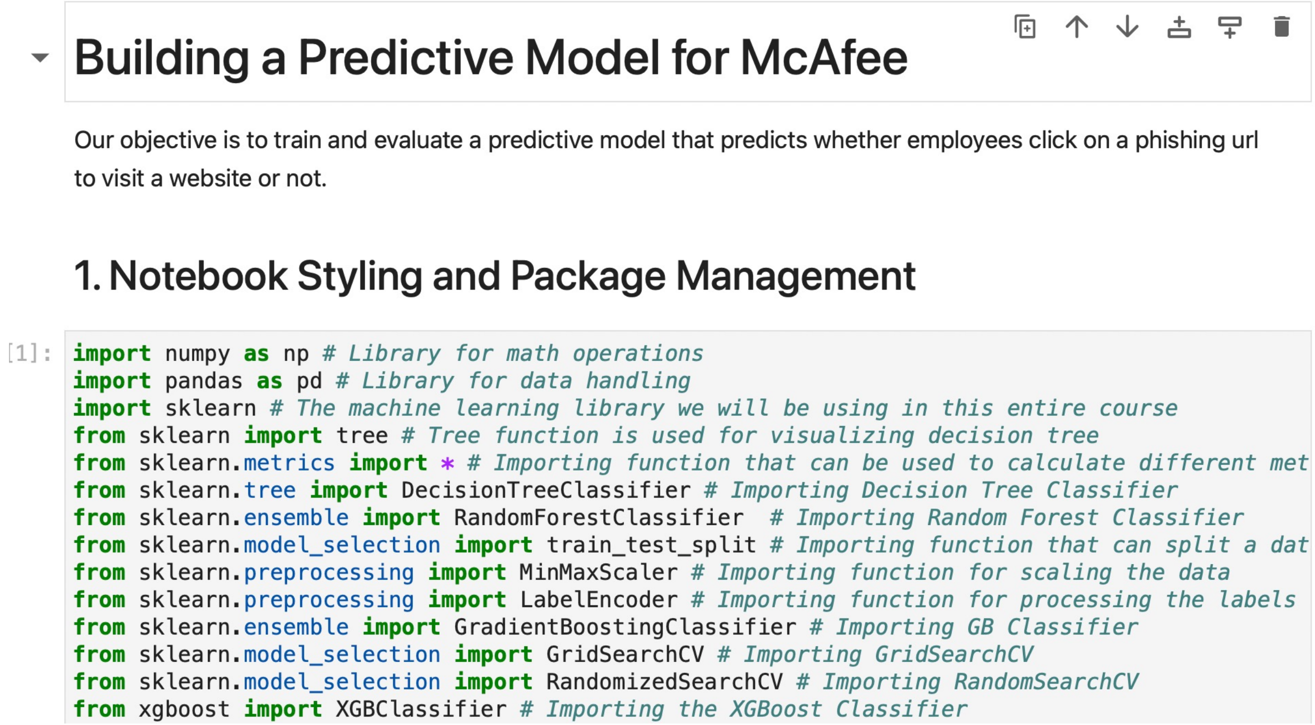
- Go to Module 2 and click on McAfee_PredictiveModel.ipynb

- In your Mod 3 teams, build the best predictive model you can.

- Submit your solutions to Kaggle.

- Be ready to discuss your approach for building your best model.

The Notebook



Building a Predictive Model for McAfee

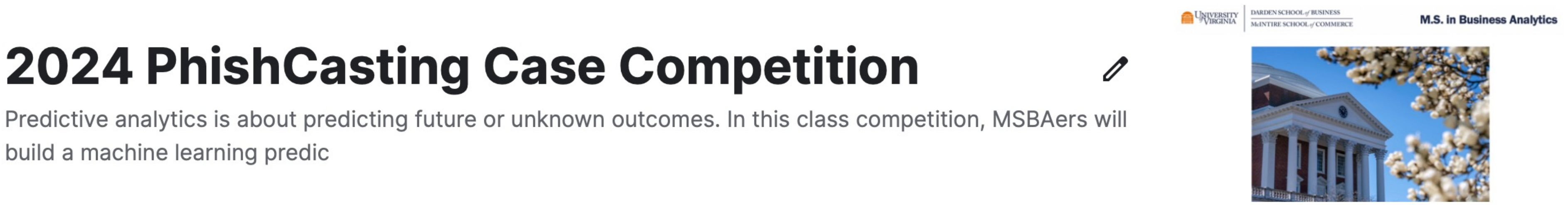
Our objective is to train and evaluate a predictive model that predicts whether employees click on a phishing url to visit a website or not.

1. Notebook Styling and Package Management

[1]:

```
import numpy as np # Library for math operations
import pandas as pd # Library for data handling
import sklearn # The machine learning library we will be using in this entire course
from sklearn import tree # Tree function is used for visualizing decision tree
from sklearn.metrics import * # Importing function that can be used to calculate different met
from sklearn.tree import DecisionTreeClassifier # Importing Decision Tree Classifier
from sklearn.ensemble import RandomForestClassifier  # Importing Random Forest Classifier
from sklearn.model_selection import train_test_split # Importing function that can split a dat
from sklearn.preprocessing import MinMaxScaler # Importing function for scaling the data
from sklearn.preprocessing import LabelEncoder # Importing function for processing the labels
from sklearn.ensemble import GradientBoostingClassifier # Importing GB Classifier
from sklearn.model_selection import GridSearchCV # Importing GridSearchCV
from sklearn.model_selection import RandomizedSearchCV # Importing RandomSearchCV
from xgboost import XGBClassifier # Importing the XGBoost Classifier
```

Kaggle Competition



2024 PhishCasting Case Competition

Predictive analytics is about predicting future or unknown outcomes. In this class competition, MSBAers will build a machine learning predic

University of Virginia
Darden School of Business
McIntire School of Commerce
M.S. in Business Analytics

After the Competition

- What was your highest AUC in Kaggle?
- What algorithm (e.g., XGBoost, RandomForest, …) gave you the best AUC?
- What hyperparameters did you adjust? What values you selected for the hyperparameters? How did you come up with these values?
- Any best practices you would like to share with the class?

?

VIRGINIA