

INFOFISH

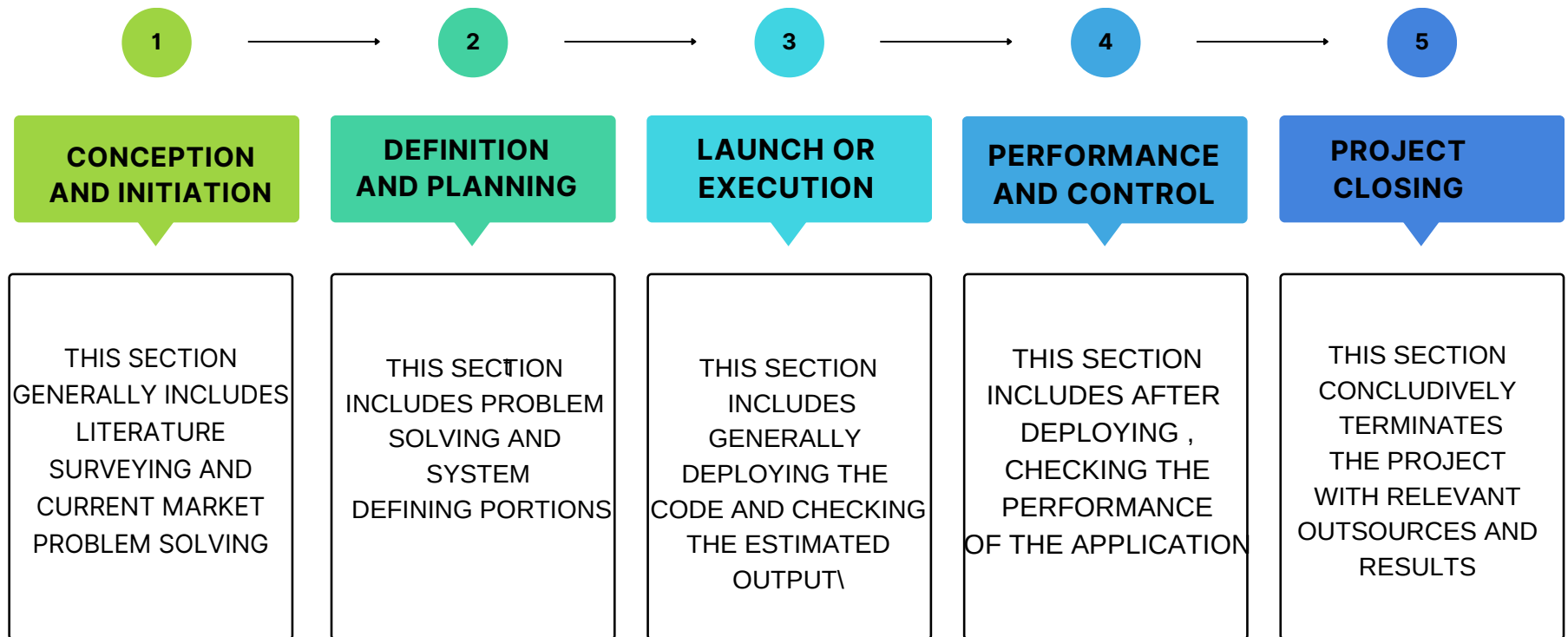
-BY ADAN BARI
03/06/2023

A TECH APP FOR DETECTION OF PHISHING WEBSITES

There's a lot of automation that can happen that isn't a replacement of humans, but of mind-numbing behaviour.

~ Stewart Butterfield

BASIC DESIGNING STEPS





STEP BY STEP BUSINESS NEED ASSESSMENT

Market Research

Market research is useful to determine the opportunity and how big a business opportunity you will build i.e. extinction of it

01

Market
Research

02

Business
Response

Business Response

Evaluates your overall business and staff's responsiveness to a phishing attempt

Fund Your Business

Determine how much capital you need and how long you can return the capital according to your business plan

03

Fund Your
Business

04

Business
Location

Business Location

The location of your business is also very influential on sales, make sure the location is strategic and safe



TARGET SPECIFICATIONS AND CUSTOMER CHARACTERIZATION

CHARACTERISTICS OF PHISHING

A web page is a document created with Hypertext Markup Language (HTML) and Cascading Style Sheets (CSS) that is displayed in a web browser. The address bar of a browser specifies the URL of the website. A website is a single domain name that consists of a collection of interlinked web pages. The characteristics of a phishing website were examined in this chapter. A web page is made up of three parts: HTML, CSS, and JavaScript. Attackers use these sections of code to carry out a wide range of phishing schemes.

200+

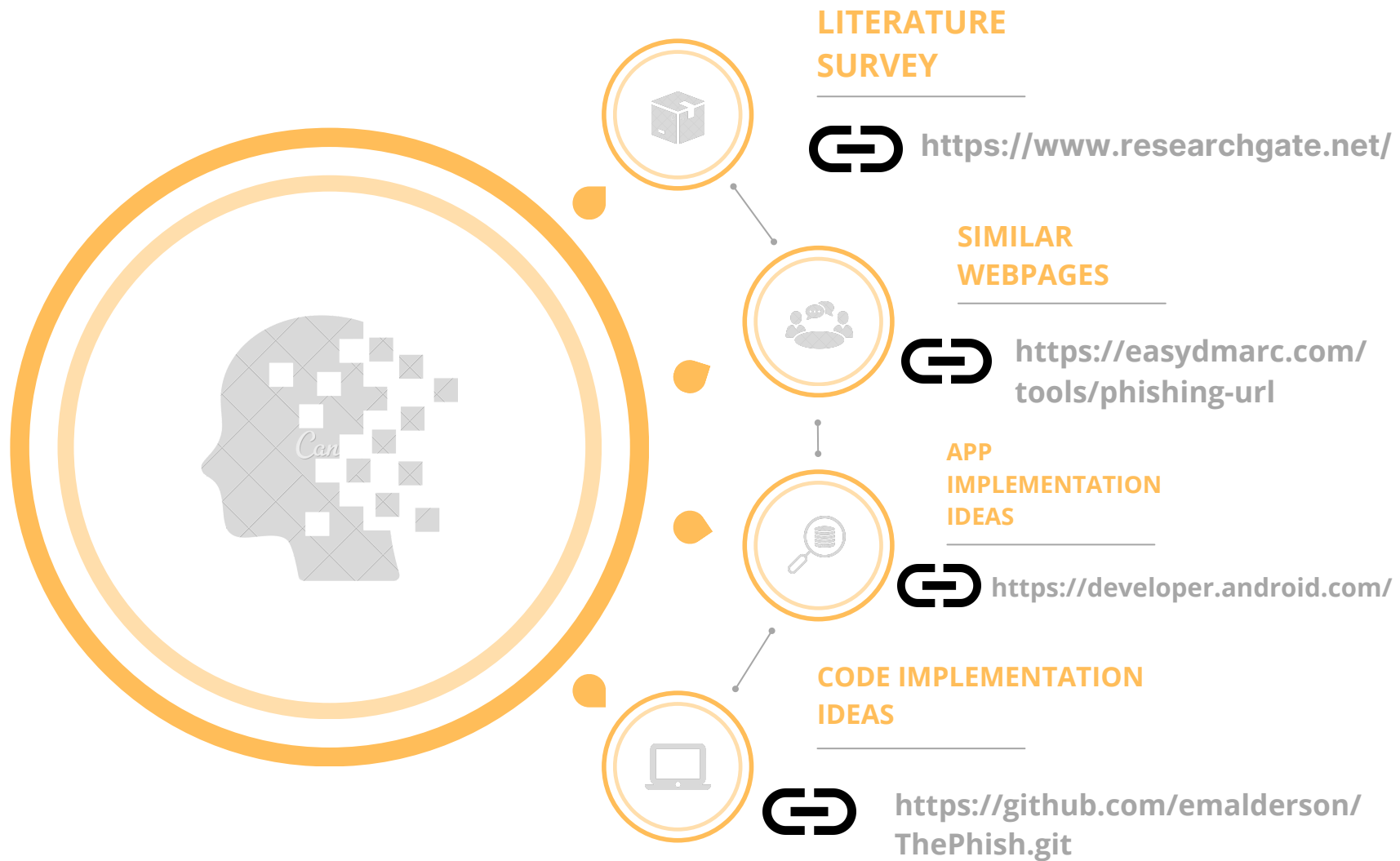
Start a branding
Business processes
should have clear
objectives

117+

Begin new ideas
A visual representation
of the process model
should be created



EXTERNAL REFERENCE



BENCHMARKING ALTERNATIVE PRODUCTS

PHISHER HUNTER



ABDULLAH ALGUMAIJAIN

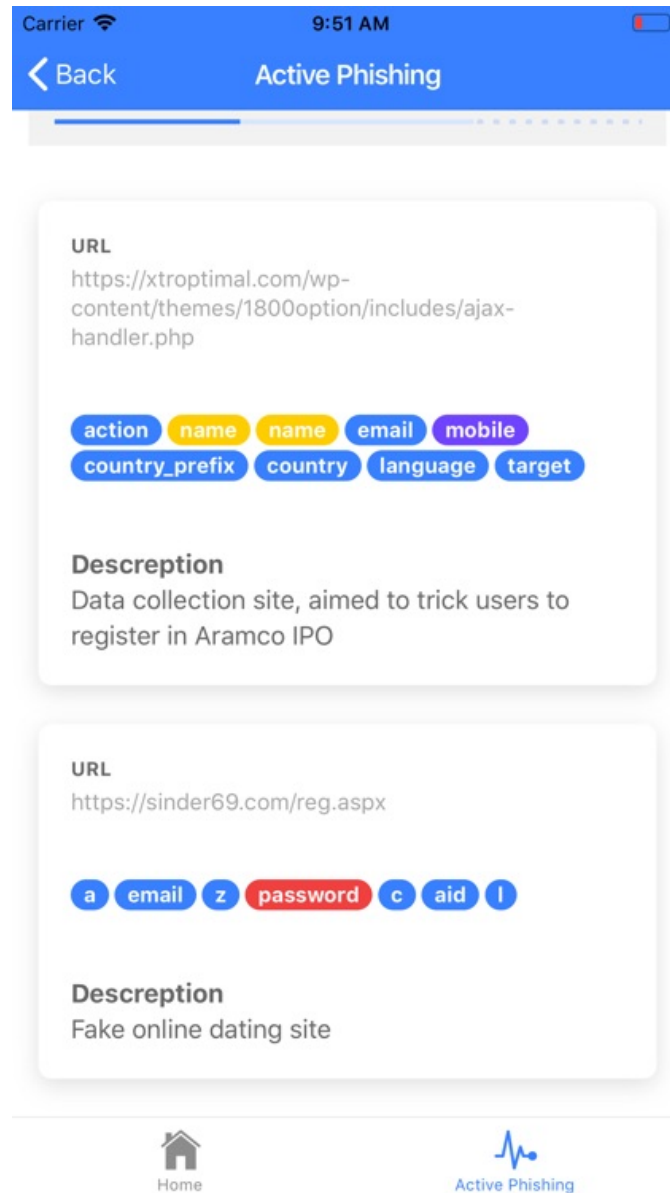


COPYRIGHTS TO DEVELOPER OF THIS APP



PHISHER HUNTER INC.: IS A USER INTERFACE APPLICATION DEVELOPED BY ABDULLAH ALGUMAIJAIN AVAILABLE AT VARIOUS DOWNLOADING PLATFORMS i,e APP STORE,PLAY STORE etc; WHICH OFFERS VARIOUS FEATURES SUCH AS PHISHER GENERATOR FOR IDEA TOUCHUP AND PROVIDES TRACKING,DESCRIPTION AND REPORTING OPTIONS ADDITIONALLY.

USER INTERFACE OF THE APP



MONETIZATION MODEL

BUSINESS MODELLING

A business model is nothing more than a blueprint that explains how a company aims to make money. It explains who your customer base is, how you provide value to them, and the financial numbers that go along with it. The business model canvas allows you to define all of these elements on a single page. A business model is nothing more



MODEL DESCRIPTION



THE FREE AND PAID APP VERSIONS MODEL

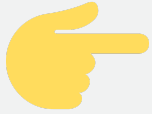
ONE POPULAR APP MONETIZATION STRATEGY IS TO OFFER BOTH FREE AND PAID VERSIONS OF YOUR APP. WITH THIS APPROACH APP DEVELOPERS WILL EITHER LIMIT CERTAIN FEATURES IN THE FREE APP IN ORDER TO "ENCOURAGE" THE FREE APP USER TO UPGRADE PAID APP, OR MONETIZE THE FREE APP WITH IN-APP ADVERTISING.

THE BENEFITS WITH THIS STRATEGY IS TWO FOLD. ON ONE HAND IT PROVIDES A FREE OPTION FOR USERS TO EXPERIENCE THESE FUNCTIONALITY OF APP AT NO COST. WHILE ON THE OTHER HAND, DEVELOPER WITH A GROWING USER BASED POTENTIAL VIA APP UPGRADES AND ADVERTISING



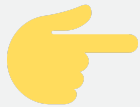
THE FREE APP WITH IN -APP PURCHASES MODEL

ANOTHER POPULAR APP MONETIZATION STRATEGY IS IN-APP PURCHASES OR IAP. THE APP ITSELF AND USUALLY GIVES THE BASIC FEATURES OF THE APP, ARE FREE. HOWEVER, IF THE USER WOULD LIKE TO ADVANCE IN THE APP FASTER, SAY GET EXTRA BENEFITS THERE ON



THE PAID APP MODEL

IN SOME CASES, DEVELOPERS WILL MAKE THEIR APPS AVAILABLE ONLY IN THE PAID VERSIONS. USUALLY APP OFFERS UNIQUE VALUE THAT'S HARD TO FIND ELSEWHERE.



THE PARTNERSHIP MODEL

IF AN APP IS VERY POPULAR AMONG A CERTAIN NICHE MARKET, COMPANIES IN THAT SPACE MAY APPROACH THE APP DEVELOPER FOR SPONSORSHIP OPPORTUNITIES TO GET THEIR BRAND IN FRONT OF YOUR APP USERS.

SO WHEN YOU'RE PLANNING ABOUT THE NEXT APP TO BUILD, YOU CAN ALSO THINK ABOUT THE TARGET USER BASE TO ATTRACT AND THE TYPE OF BRANDS THAT MIGHT BE INTERESTED IN THAT AUDIENCE.

CONCEPT GENERATION

THE IDEA OF CONCEPT CAME INTO MY MIND AFTER DEALING AND GETTING IRRITATED WITH MULTIPLE PHISHING WEBSITE WHILE DOING MY NORMAL BROWSING OR USAGE TIMING.

THE IDEA BASICALLY DREW MY ATTENTION TO KNOW ABOUT THE ENTIRE PROCESS AND IT'S ANALOGIES.MEANWHILE I WAS READING AND ANALYZING IT, I THOUGHT OF WORKING ON SOME SIMILAR CONTENT.i,e, TO DEVELOP AN APP WHICH BASICALLY PROVIDES A NORMAL SEARCH BOX OPTIMIZATION IN THE USER INTERFACE LEVEL WHICH BASICALLY CAPTURES YOUR URL //https: AND THEN CONNECT IT WITH DOMAIN TO TARGET AND PHISH A DOMAIN EXACTLY SAME AS AN ORIGINAL DOMAIN AND TELL US IN FEW SECONDS WHETHER THE URL WE ARE WORKING WITH IS LEGITIMATE OR A PHISHEDONE.

EVEN INIDA IS FACING A HUGE INCREASE IN PHISHING CRIMES YEAR AFTER YEAR, THE RESOURCES WHICH SHOULD BE IMPLEMENTED TO GET THE BEST OUT OF IT ARE:

Real-Time URL Scanning: The app scans URLs in real-time, comparing them against a comprehensive database of known phishing websites. It uses machine learning algorithms to analyze various factors as domain name, SSL certificate, page content, and website reputation to determine the likelihood of website being a phishing site.

Browser Extension Integration: The app can integrate with popular web browsers as a browser extension. When users navigate to a website, the app automatically scans the URL and displays a warning if it detects potential phishing indicators. This provides users with instant feedback and helps them make informed decisions about the websites visit.

Reporting Phishing Websites: Users can report suspicious websites directly through the app. This crowdsourced approach helps to continuously update and improve the database of known phishing websites, ensuring the app stays up-to-date and effective in detecting new phishing techniques.

User Feedback and Ratings: The app allows users to provide feedback and ratings for websites they have encountered. This information helps build a community-driven platform, where users can share their experiences and warn others about potential phishing threats.

Anti-Phishing Education: The app includes educational resources on how to identify and avoid phishing attacks. It provides tips and best practices for recognizing suspicious URLs, email phishing attempts, and other common techniques used by cybercriminals. This empowers users to become more vigilant and educated about online security.

>**Notifications and Alerts:** The app sends push notifications or email alerts to users when they encounter potentially malicious website or receive a suspicious email. This proactive approach helps users take immediate action and avoid falling victim to phishing attacks.

>**Offline Mode:** The app includes an offline mode where it can still detect phishing websites using a local database. This feature is useful when users are in areas with limited or no internet connectivity

>**Integration with Password Managers:** The app can integrate with popular password manager applications, allowing users to store their login credentials securely. It can automatically detect and block phishing attempts when users try to enter their sensitive information on fraudulent websites

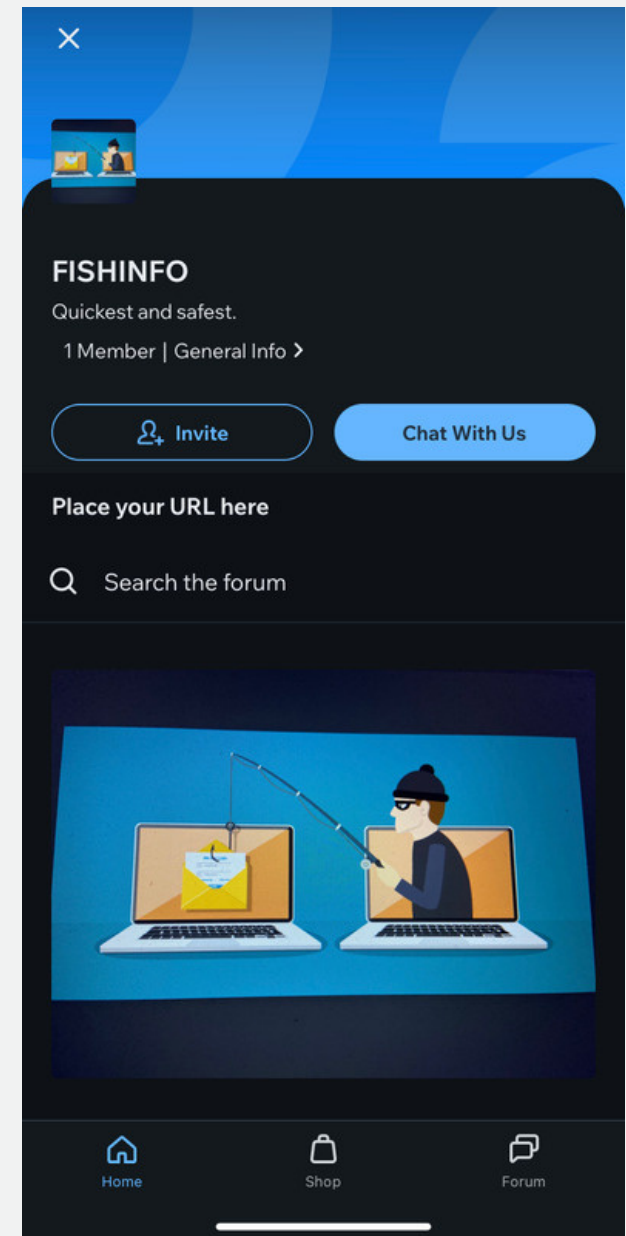
>**Multi-Platform Support:** The app is available on various platforms, including mobile devices (iOS and Android) desktop computers (Windows and macOS), and web browsers. This ensures broad accessibility and, coverage for users across different devices.

>**Privacy and Data Security:** The app prioritizes user privacy and data security. It adheres to strict privacy standards, securely handles user data, and ensures that user information is not compromised or misused in any way.

FINAL PROTOTYPE

BASIC LAYOUT OF THE APP

WITH THE HELP OF THIS APP WE CAN EASILY DETECT THE FALLACY IN THE URL OR IF IN CASE IT'S A FAKE PHISHING WEBSITE, WITH THE HELP OF MACHINE LEARNING AND SEMANTIC ANALYSIS, WE CAN CLASSIFY WHETHER A WEBSITE IS CLEAN AND SAFE OR A PHISHED ONE.



MOBILE VIEW WINDOW

STEPS OF DETECTION

AFTER PASTING THE URL IN THE APP SEARCH ENGINE BOX WE CAN ANAL AND WAIT FOR THE RESULTS TO BE DECLARED ON THE BASIS OF MACHI LANGUAGE MODELLING AND SEVERAL MACHINE METHODOLOGIES WE CA PREDICT WHETHER A WEBSITE IS GENUINE OR NOT?

The image displays a web security analysis tool interface. At the top, there is a search bar with the placeholder text "Enter a URL: www.example.com" and a blue "SCAN" button. Below the search bar, the URL "https://www.amazon.in/gp..." is entered, with buttons for "Clean" and "Dispute". To the right of the URL are buttons for "Pivot", "DOM", and "VIEW WHOIS INFO".

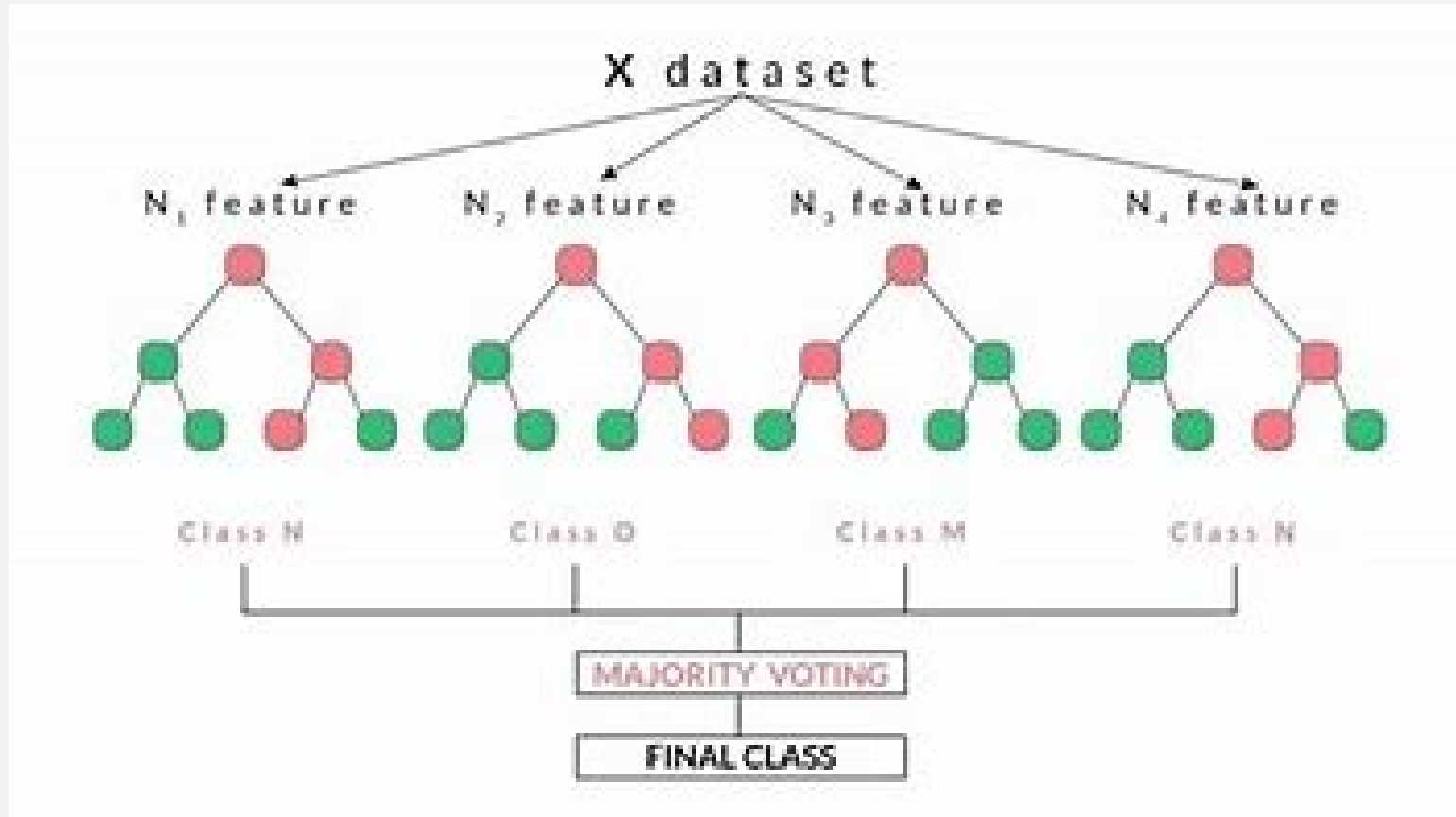
The main section is titled "Scan Results" and contains the following information:

- Source URL:** https://www.amazon.in/gp/help/customer/display.html?nodeId=...
- Brand:** Unknown
- Redirected URL:** https://www.amazon.in/gp/help/customer/display.html?nodeId=...
- IP Address:** 54.192.74.77
- Hosting Provider:** Amazon.com, Inc.
- First Seen:** June 6th 2023, 12:59:57 am
- Category:** Gift Card
- ASN:** 16509
- Location:** United States
- Abuse Contact:** abuse@akamai.com
- Certificate Details:** DigiCert Inc: www.amazon.co.in, www.amazon.in, amazon.co.in, amazon.in, origin-www.amazon.in, p-nf-www-amazon-in-kallias.a...

At the bottom left, there is a section titled "Threat Intelligence".

On the right side, there is a "Screenshot" section showing a screenshot of the Amazon website. The screenshot includes the Amazon logo, navigation links, and a "Help and Customer Service" section. The "Help and Customer Service" section has a search bar and a list of links: "Your Security", "Safe Online Shopping", "Do Not Share Personal Information", "Always Pay Via Amazon Marketplace", and "Check your Seller's Feedback".

MACHINE MODEL USED



RANDOM FOREST ALGORITHM

QUESTION ARISES, WHY WE HAVE USED RANDOM FOREST ONLY?

MODEL SELECTION CRITERIA

TEST ID	ML MODEL	TRAIN ACCURACY	TEST ACCURACY
TC01	http://www.paypal-account.com	0.923487	Phishing
TC02	https://www.amazon.com	0.076513	Legitimate
TC03	https://www.cricbuzz.in	0.038755	Legitimate

AS WE CAN SEE HERE WITHIN THE EXAMINED DATASET, RANDOM FOREST PRODUCES MOST ACCURATE PREDICTIONS WITH LESS TIME CONSUMPTION

SOURCE CODE SNIPPETS

```
from google.colab import drive
drive.mount('/content/drive')
```

Python

```
# Importing necessary libraries
```

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

Python

```
# Phishing is a type of semantic attack,4 often used to steal user sensitive information including login credentials and
# credit card numbers.5-7 It occurs when an attacker, masquerading as a trusted entity, entice a victim into clicking on a link
# or opening an attachment in an email or instant message through social messaging services such as WhatsApp, Viber or Facebook Messenger.
# The data that we have below is representing the URL features.
```

Python

```
fishing = pd.read_csv("drive/MyDrive/Dataset/Phishing.csv")
fishing.head(5)
```

Python

```
fishing = fishing.drop(['id'], axis = 1)
```

Python

```
# Welp, this is a very big dataset that has alot of information about webistes, which aslo includes their meta-data, last column
# depicts what we are trying to predict, if the website is a phishing site or not.
```

Python

```
fishing.shape
```

Python

```
fishing.shape
```

Python

```
# But before moving forward, can somebody realize what problems will arise due to such a dimensional data?  
# Would you fill a form or provide 50 such information just to see if the site is a phishing site or not?
```

Python

```
fishing.isnull().sum()
```

Python

+ Code + Markdown

```
# This dataset contains 48 features extracted from 5000 phishing webpages and 5000 legitimate webpages.
```

```
fishing.describe()
```

Python

```
# Classes to predict from.
```

```
fishing['CLASS_LABEL'].value_counts()
```

Python

```
# Finding more from the data
```

```
fishing_class = fishing.groupby('CLASS_LABEL')  
fishing_class['NoHttps'].value_counts()
```

Python

```
fishing_class['UrlLength'].mean()
```

Python

```
fishing_class['NumPercent'].mean()
```

Python

```
fishing_class['NumAmpersand'].mean()
```

Ln 2, Col 1 Spans: 4 CRLF Cell 1 of 29 Go Live

GOGGLE COLAB

```
fishing_class['IpAddress'].value_counts()
```

Python

```
plt.figure(figsize=(30, 30))
sns.heatmap(fishing.corr(),annot=True,cmap='viridis',linewidths=.5)
```

Python

```
fishing.columns
```

Python

```
subset_fishing = fishing[['NumDots', 'PathLevel', 'NumDash', 'NumSensitiveWords', 'PctExtHyperlinks',
                          'PctExtResourceUrls', 'InsecureForms', 'PctNullSelfRedirectHyperlinks', 'FrequentDomainNameMismatch',
                          'SubmitInfoToEmail', 'IframeOrFrame', 'CLASS_LABEL']]

subset_fishing.head()
```

Python

```
subset_fishing.shape
```

Python

```
y = subset_fishing['CLASS_LABEL']
X = subset_fishing.drop(['CLASS_LABEL'], axis = 1)
```

Python



```
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.ensemble import RandomForestClassifier
```

Python

```
Xtrain, Xtest, ytrain, ytest = train_test_split(X, y, test_size=0.2, random_state = 42)
```

Python

```
random_model = RandomForestClassifier(n_estimators=250, n_jobs = -1)
```

In 2: Col 1 Spaces: 4 CRLF Cell 1 of 29 Go Live  

```
#Fit
random_model.fit(Xtrain, ytrain)

y_pred = random_model.predict(Xtest)

#Checking the accuracy
random_model_accuracy = round(random_model.score(Xtrain, ytrain)*100,2)
print(round(random_model_accuracy, 2), '%')
```

Python

```
random_model_accuracy1 = round(random_model.score(Xtest, ytest)*100,2)
print(round(random_model_accuracy1, 2), '%')
```

Python

```
# Save the trained model as a pickle string.
import pickle

saved_model = pickle.dump(random_model, open('drive/MyDrive/Dataset/Models/Phishing.pickle','wb'))
```

Python

Python

TECHNICAL SPECIFICATIONS AND TEAM REQUIREMENTS

Project Manager: Oversees the entire development process, ensures the project stays on track manages timelines, and coordinates team members.

Business Analyst: Gathers and analyzes requirements, defines the scope of the app, conducts market research, and identifies target users and their needs.

UI/UX Designer: Designs the user interface (UI) and user experience (UX) of the app, focusing on creating an intuitive and visually appealing interface that enhances usability and user engagement.

App developer: To design and give layouts of the app, by designing the app interface.

Machine Learning Engineer: If you plan to incorporate machine learning algorithms or artificial intelligence techniques into the app's detection mechanisms, a machine learning engineer can help develop and fine-tune those models.

Database Administrator (DBA): Manages the database infrastructure, including data storage, retrieval, and optimization for efficient and secure access.

DevOps Engineer: Sets up the deployment infrastructure, implements continuous integration and delivery (CI/CD) pipelines, and ensures smooth deployment and monitoring of the app

REFERENCES

1. Anti-Phishing Working Group (APWG): The APWG is an international coalition that provides resources, research, and industry best practices to combat phishing attacks. Their website contains valuable information on phishing detection and prevention techniques.
 - Website: <https://apwg.org>
1. OWASP: The Open Web Application Security Project (OWASP) is a community driven organization focused on improving software security. They offer various resources related to web application security, including guidance on phishing prevention and detection.
2. Phish'Tank: PhishTank is a community-driven database of known phishing URLs. You can access their API to check URLs against their database and verify if they are phishing attempts.
 - Website: <https://www.phishtank.com/>