IBM Z Datathon

PhishTank

THEME: TECHNOLOGY FOR GOOD

TEAM CODE RED

Thirumurugan RA - Team Leader & LLM Engineer

Sudarshan Olirvel - Data Scientist and Solution Architect

Ananya Raman - Frontend Developer and Research

Shrika Thota - Data Engineer and Backend Integration

Madussree Ravi - Backend Developer and Documentation

PhishTank - Real-Time Phishing Detection & Protection

Phishing: *The Modern Threat Vector*

- *Phishing*: A cyber attack using digital deception to steal critical data like credentials and financial information.
- *The Modern Threat:* Attacks are now AI-powered and highly targeted, designed to bypass traditional security filters.
- *Conventional Defenses are Obsolete*: Reactive tools like spam filters and antivirus are insufficient against sophisticated, zero-day threats.
- ***Our Solution:*** proactive, AI-driven defense that analyzes and neutralizes threats before they reach the user by predicting deceptive behavior.

PhishTank

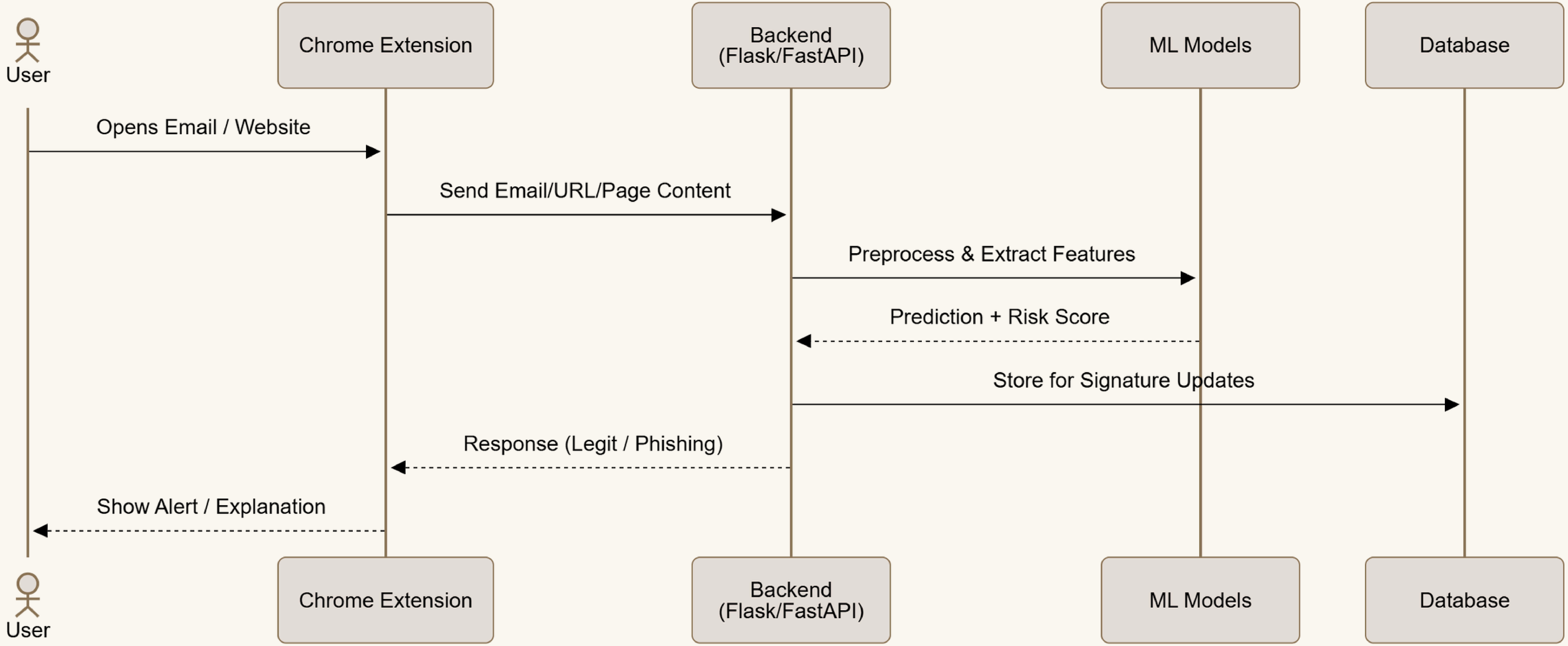
Tech for Good, Trust for All

Context Matters: ***Every Click Tells a Story.***

As cyber threats grow more *adaptive*, PhishTank turns data into defense - why not beat attackers in their own race?

From suspicious URLs to deceptive emails, we empower users to stay ahead of digital deception with accuracy, speed, and clarity.

Real-Time Detection Pipeline



Datasets for Model Training & Evaluation

PhishTank encompasses two **pipelines** — one for URL-based phishing detection and another for email-based phishing detection.

URL Dataset

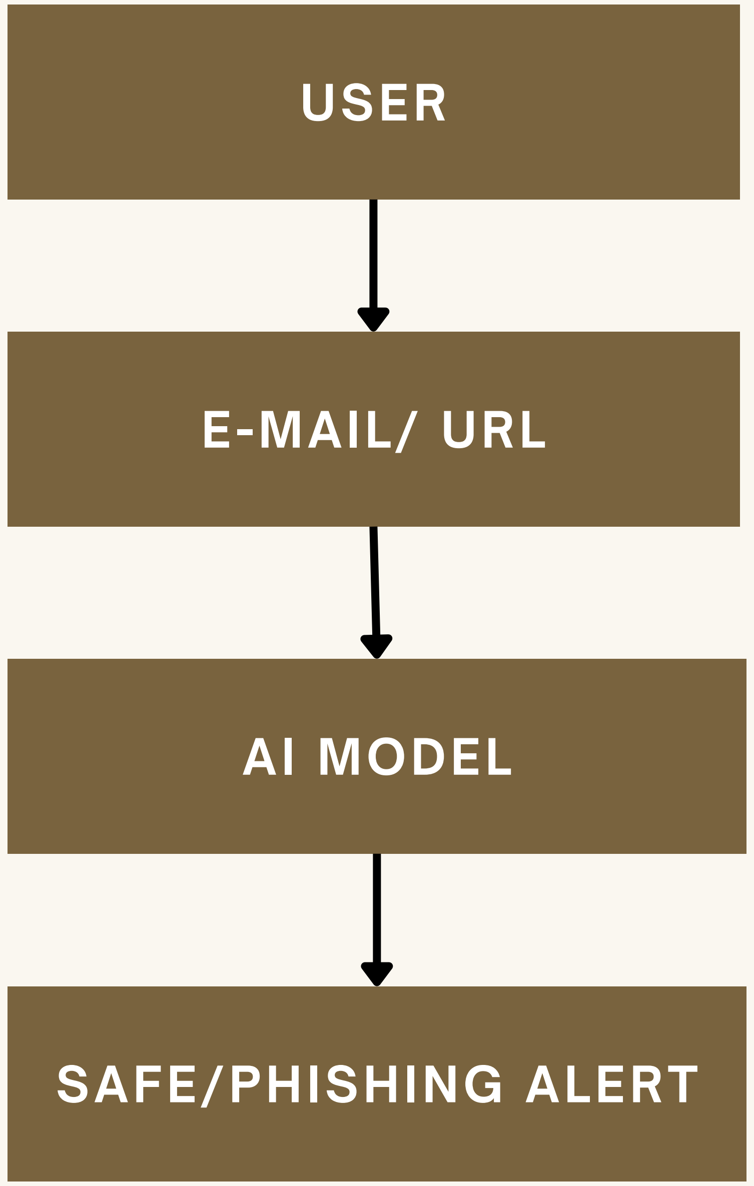
- Source: *Web Page Phishing Detection Dataset – Kaggle*
- Description: Contains labeled phishing and legitimate website URLs with extracted lexical and structural features such as URL length, number of dots, HTTPS presence, and domain-related attributes.

Email Dataset

- Source: Phishing Email Dataset – Zenodo
- Description: Contains full text, subjects, and metadata enabling learning linguistic and contextual phishing indicators such as urgency cues, spoofed brand names, and credential requests.

Labels:

- *0 → Legitimate*
- *1 → Phishing*



USER

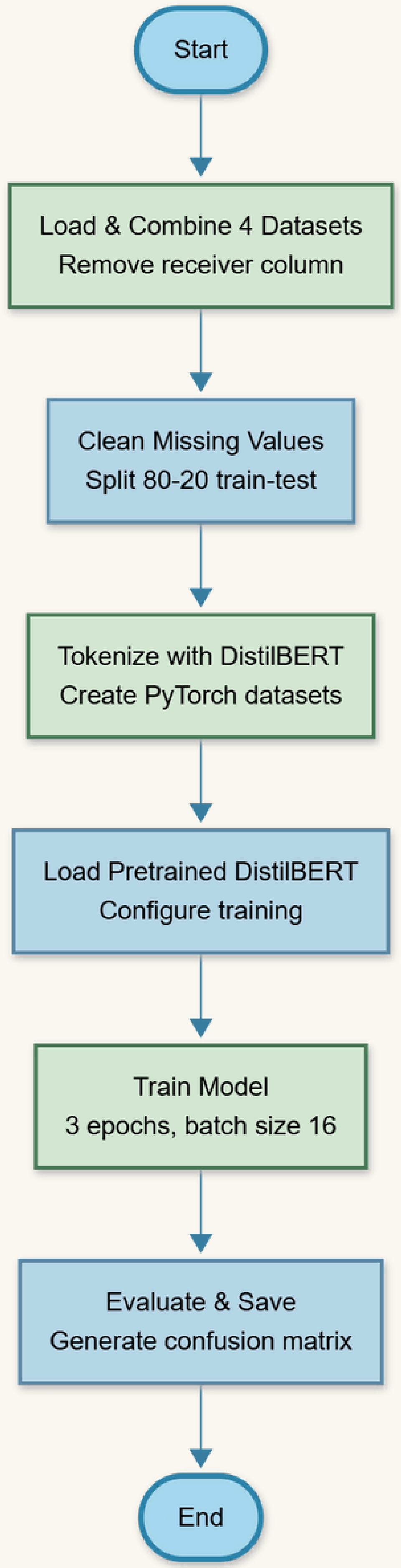
E-MAIL/ URL

AI MODEL

SAFE/PHISHING ALERT

Email Phishing Detection

TEAM CODE RED



Start

Load & Combine 4 Datasets
Remove receiver column

Clean Missing Values
Split 80-20 train-test

Tokenize with DistilBERT
Create PyTorch datasets

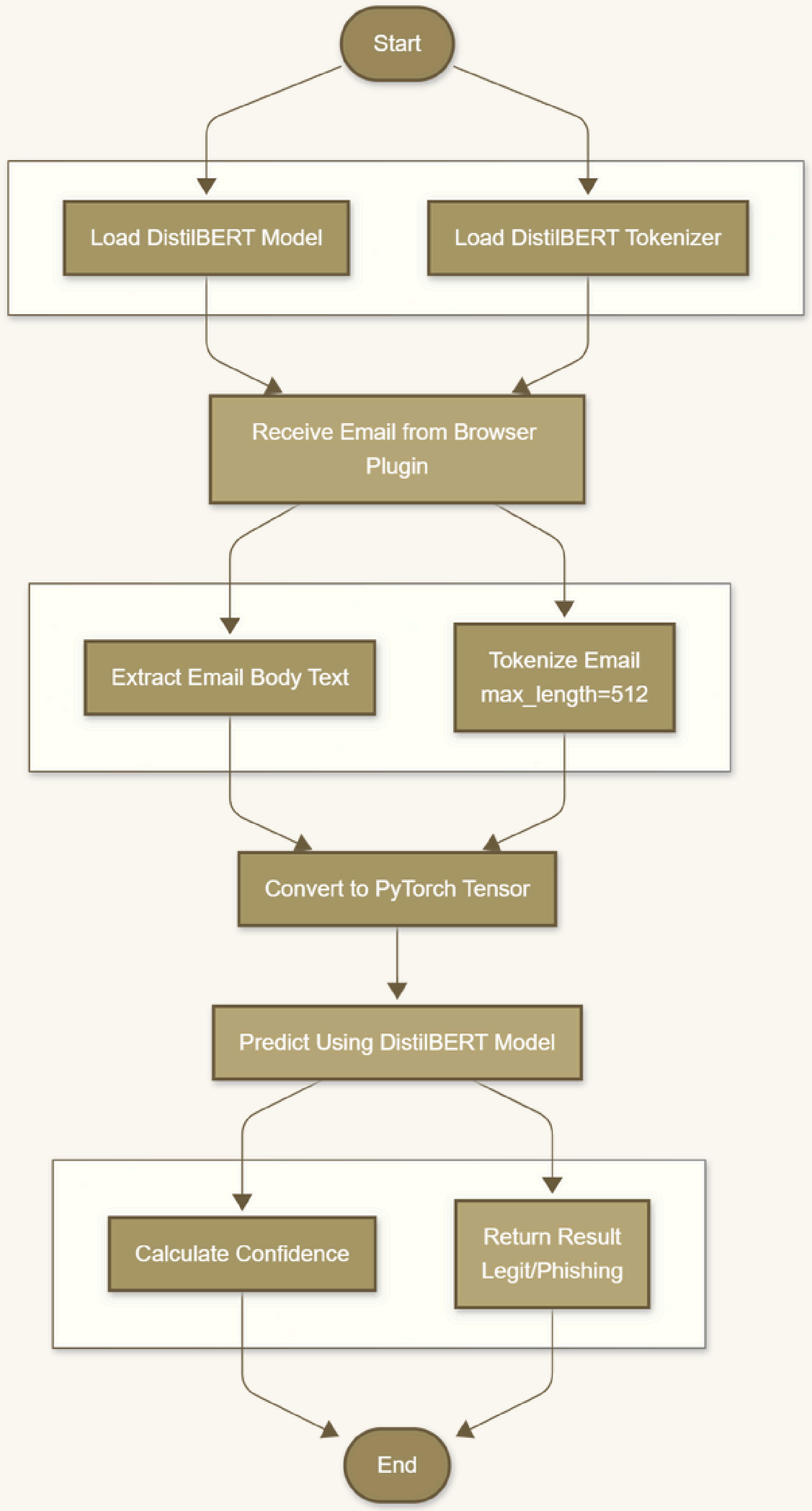
Load Pretrained DistilBERT
Configure training

Train Model
3 epochs, batch size 16

Evaluate & Save
Generate confusion matrix

End

Model Training Pipeline



Start

Load DistilBERT Model

Load DistilBERT Tokenizer

Receive Email from Browser Plugin

Extract Email Body Text

Tokenize Email
max_length=512

Convert to PyTorch Tensor

Predict Using DistilBERT Model

Calculate Confidence

Return Result
Legit/Phishing

End

Data Flow Pipeline

Model: DistilBERT

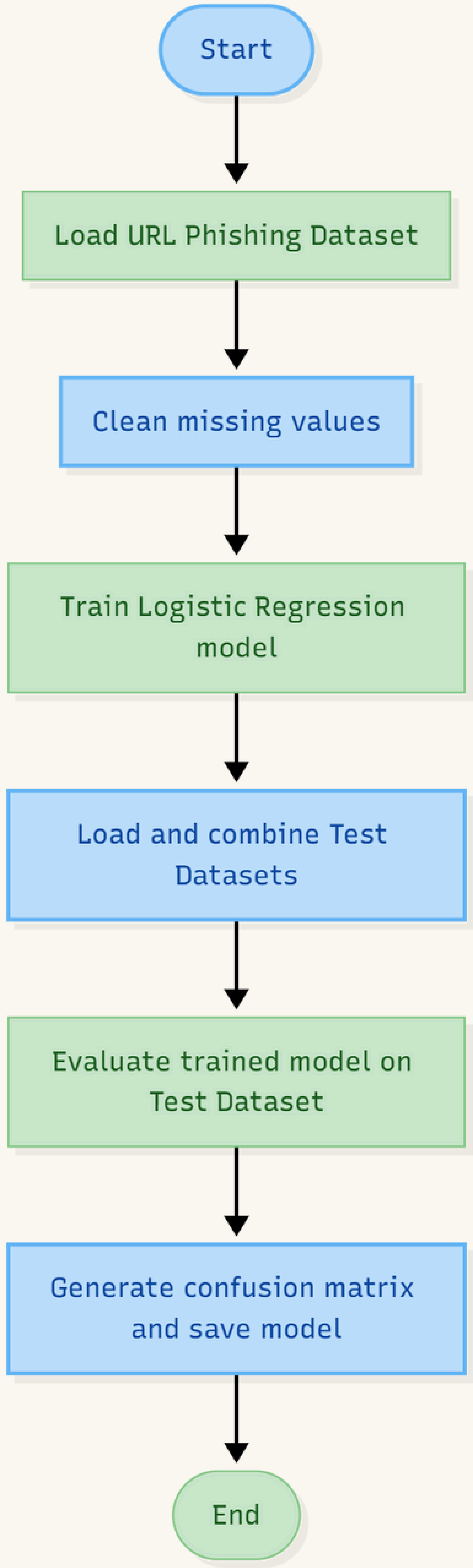
- Why - Delivers BERT-level contextual understanding while being 40% smaller and 60% faster, enabling efficient phishing email classification.

Implementation Highlights

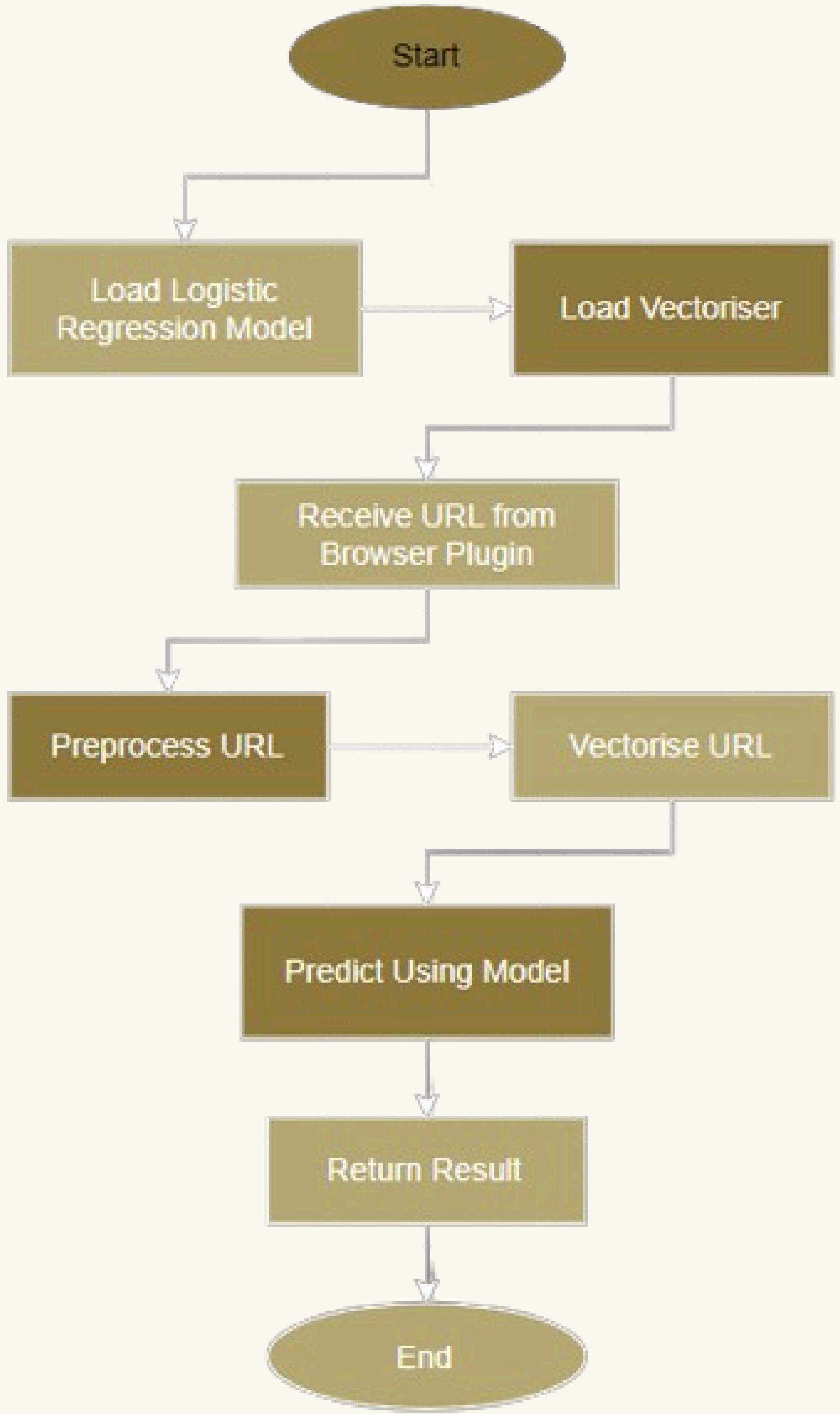
- Training Data: 4 combined datasets with 30,000+ emails
- Performance: 99.58% accuracy across all metrics
- Architecture: Pretrained model fine-tuned for binary classification
- Deployment: FastAPI backend with **<100ms response time**

URL Phishing Detection

TEAM CODE RED


Start
Load URL Phishing Dataset
Clean missing values
Train Logistic Regression model
Load and combine Test Datasets
Evaluate trained model on Test Dataset
Generate confusion matrix and save model
End

Model Training Pipeline



Data Flow Pipeline

Model : Logistic Regression

- Why – Utilizes vectorized URL features with logistic regression, delivering high accuracy and fast execution for real-time phishing detection through a lightweight, efficient model.

Implementation Highlights

- Training Data: 11,000+ labeled URLs for model training
- Testing Data: Combined dataset of 500K+ URLs
- Performance: 94.4% accuracy with 94–98% precision/recall
- Architecture: Logistic Regression with URL vectorization
- Deployment: FastAPI backend, <100ms response time

Tech Stack

Frontend (Browser Extension)

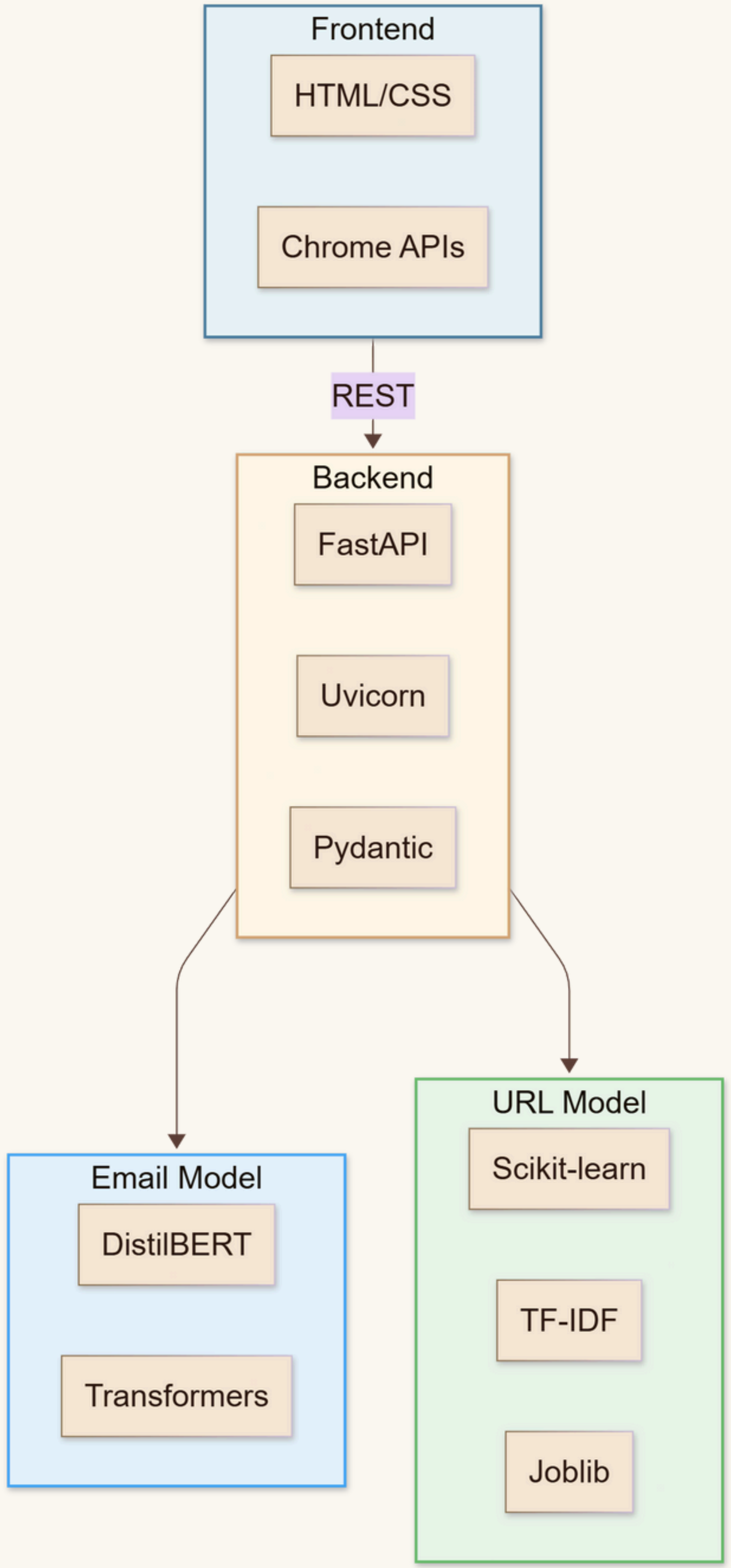
- HTML/CSS - UI design
- Chrome Extension APIs

Backend (API Server)

- FastAPI - Python web framework
- Uvicorn - ASGI server
- Pydantic - Data validation and serialization

Machine Learning Models

1. *URL Classification*
   - Scikit-learn - Logistic Regression model
   - TF-IDF Vectorizer - Feature extraction
   - Joblib - Model serialization
2. *Email Classification*
   - DistilBERT - Transformer-based NLP model
   - Hugging Face Transformers - Pre-trained model integration



TEAM CODE RED

Demo - URL pipeline

Header Section

- Navigation Tabs: Three-tab layout with "Threats", "Whitelist", and "Blacklist" sections

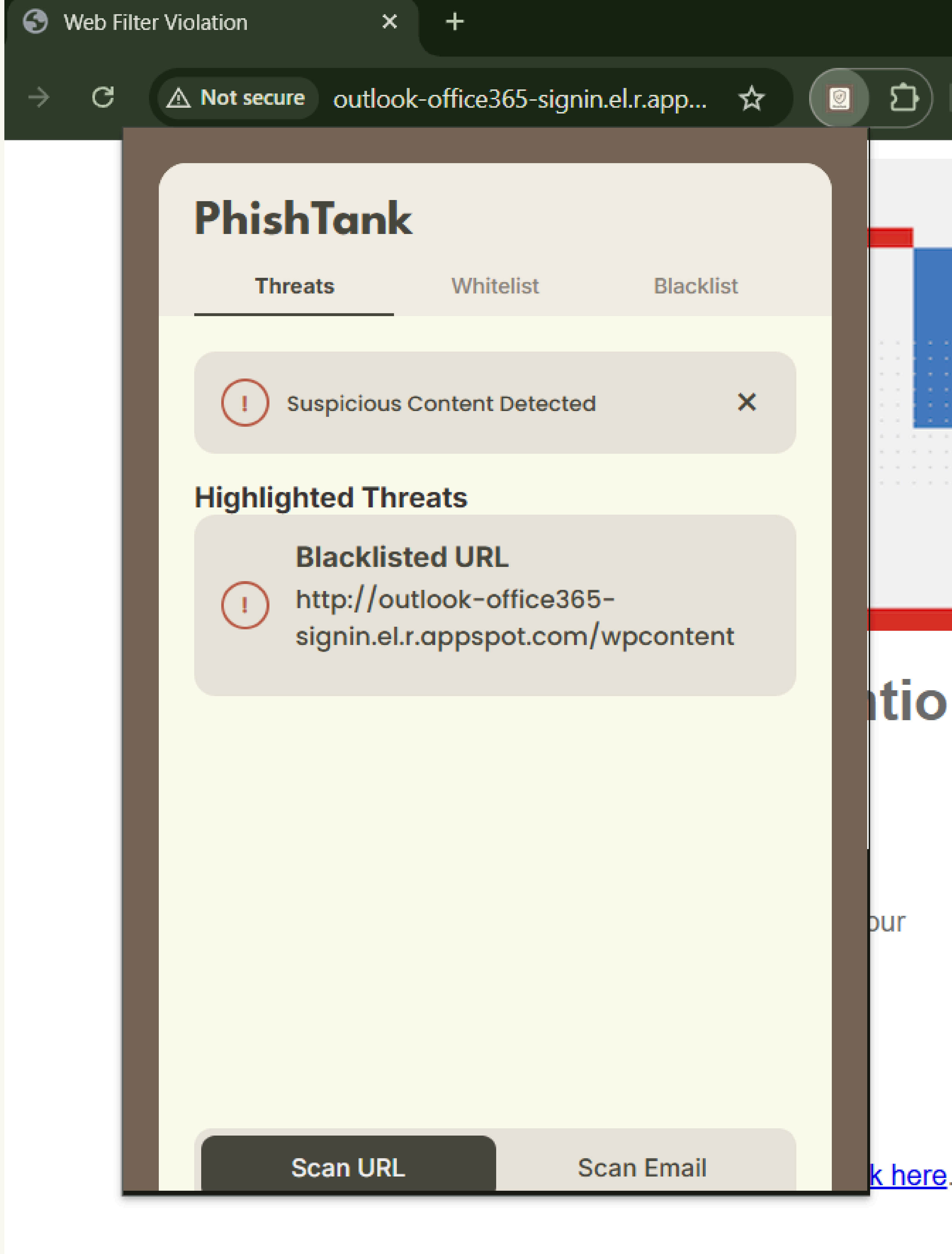
Detection Status Display

- Alert Banner + Real-time Scanning

Highlighted Threats Section

- Threat Categorization:
  - Suspicious URL
  - Whitelisted URL

- Explanations:
  - Toast-style descriptions explaining why content is flagged
  - Contextual Warnings: Specific explanations like "This looks like a spoofed login page.



Suspicious Content Detected

Highlighted Threats

Blacklisted URL
http://outlook-office365-signin.el.r.appspot.com/wpcontent

TEAM CODE RED

Demo - Email pipeline

When the "*Scan Email*" button is clicked, the email pipeline is activated

Header Section

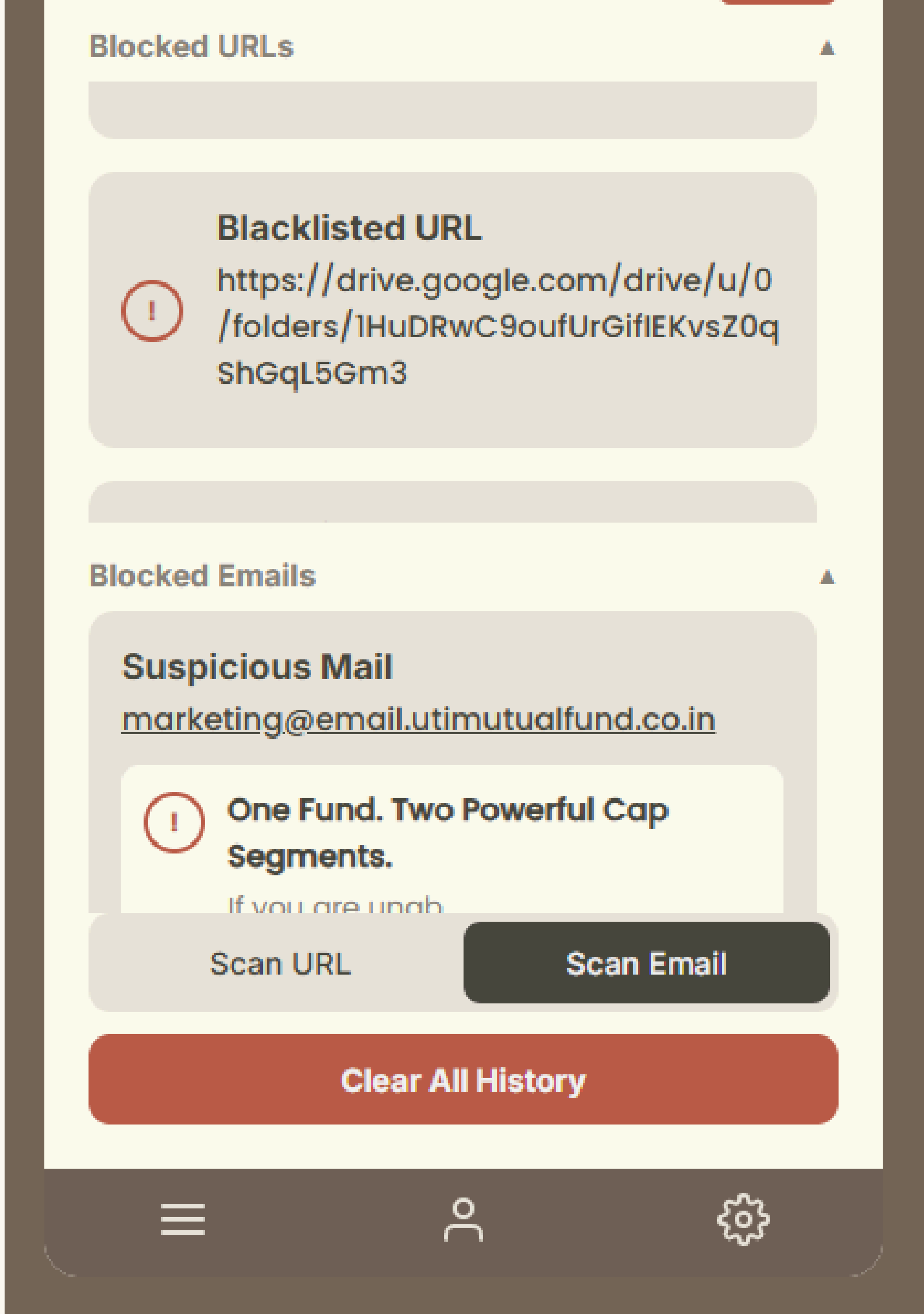
- Navigation Tabs: Two-tab layout with "Whitelist", and "Blacklist" sections

Detection Status Display

- Notification + Classification

Explanations:

  - Toast-style descriptions explaining why content is flagged
  - If **phishing email**, then it goes under the *Blacklist tab*



Blacklisted URL
https://drive.google.com/drive/u/0/folders/1HuDRwC9oufUrGifIEKvsZ0qShGqL5Gm3

Suspicious Mail
marketing@email.utimutualfund.co.in

One Fund. Two Powerful Cap Segments.
If you are unab

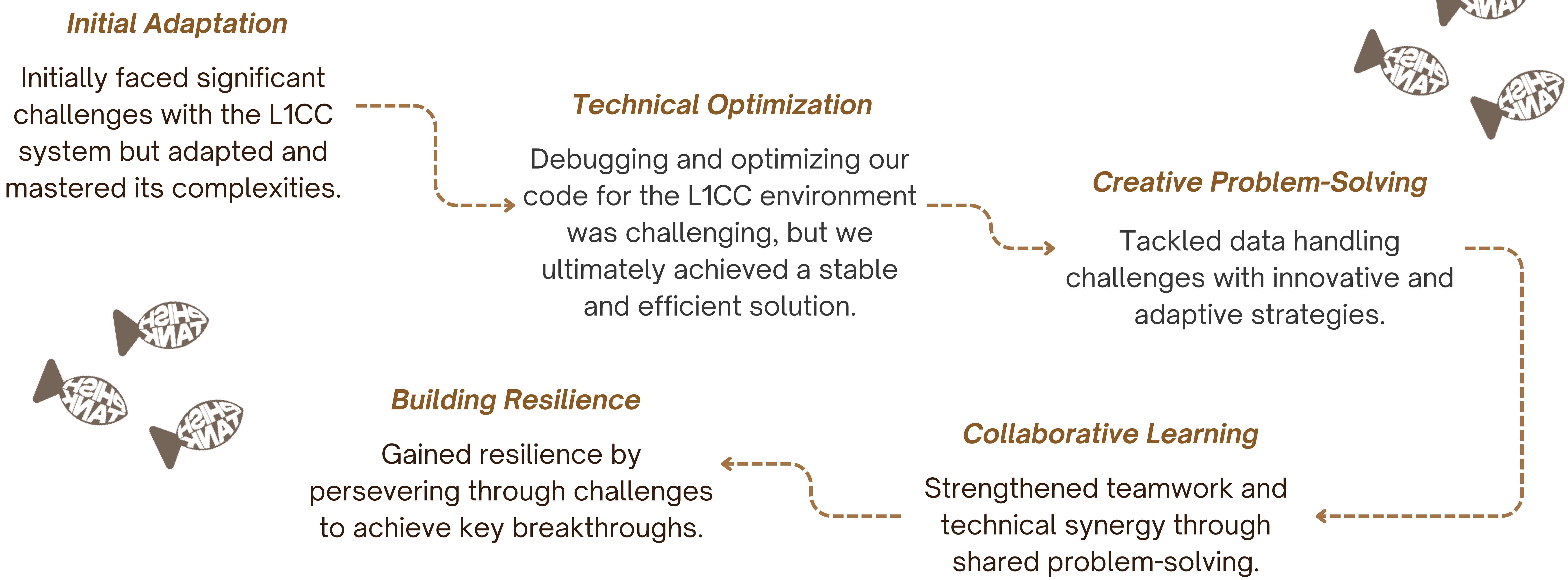
Our Differentiators

Fortifying Digital Trust & Empowering Your Workforce

PhishTank shield protects your people by understanding the language of deception.

- *Dual-Engine AI: We combine high-speed technical analysis with deep contextual understanding to stop both sophisticated exploits and psychological manipulation.*
- *Proven Accuracy: Our 99.58% accuracy transforms your greatest vulnerability into a fortified defense.*
- *Real-Time Protection: An always-on threat shield neutralizes threats seamlessly as users browse—before they can cause harm.*
- *Fostering Resilience: By eliminating the risk of human error, we enable your team to innovate securely and with confidence.*

Our Datathon Experience: A Learning Retrospective



Initial Adaptation

Initially faced significant challenges with the L1CC system but adapted and mastered its complexities.

Technical Optimization

Debugging and optimizing our code for the L1CC environment was challenging, but we ultimately achieved a stable and efficient solution.

Creative Problem-Solving

Tackled data handling challenges with innovative and adaptive strategies.

Collaborative Learning

Strengthened teamwork and technical synergy through shared problem-solving.

Building Resilience

Gained resilience by persevering through challenges to achieve key breakthroughs.