SoftDev P04: Makers Makin' It, Act II
ZealousZebras: Suhana Kumar, Anastasia Lee, Christopher Louie, Dua Baig
TARGET SHIP DATE: 2025-04-23

Overview

Our site is meant to give the user an overview of data regarding cyberattacks. This data includes: severity level, affected industries, methods of attack, dates, etc. The user can see various graphs generated based on this data, and can filter out which data they wish to see.

Dataset:

🌐 Global Cybersecurity Threats (2015-2024)

- (https://www.kaggle.com/datasets/atharvasoundankar/global-cybersecurity-threats-2015-2024/data)

Description: The Global Cybersecurity Threats Dataset (2015-2024) provides extensive data on cyberattacks, malware types, targeted industries, and affected countries. It is designed for threat intelligence analysis, cybersecurity trend forecasting, and machine learning model development to enhance global digital security.

Format: CSV
File size: 45 MB
Columns:

1. **Country**: Country where the attack occurred
2. **Year**: Year of the incident
3. **Attack Type**: Type of cybersecurity threat (e.g., Malware, DDoS)
4. **Targeted Industry**: Industry targeted (e.g., Finance, Healthcare)
5. **Financial Loss** (in million $): Estimated financial loss in millions
6. **Number of Affected Users**: Number of users impacted by the attack
7. **Attack source**: Origin of the attack (e.g., Nation-State, Hacktivist)
8. **Security Vulnerability Type**: Type of vulnerability exploited (e.g., Zero-Day, SQL Injection)
9. **Defense Mechanism Used**: Cyber defense strategy applied (e.g., Firewall, IDS)
10. **Incident Response Time** (in hours): Time taken to fully resolve the incident

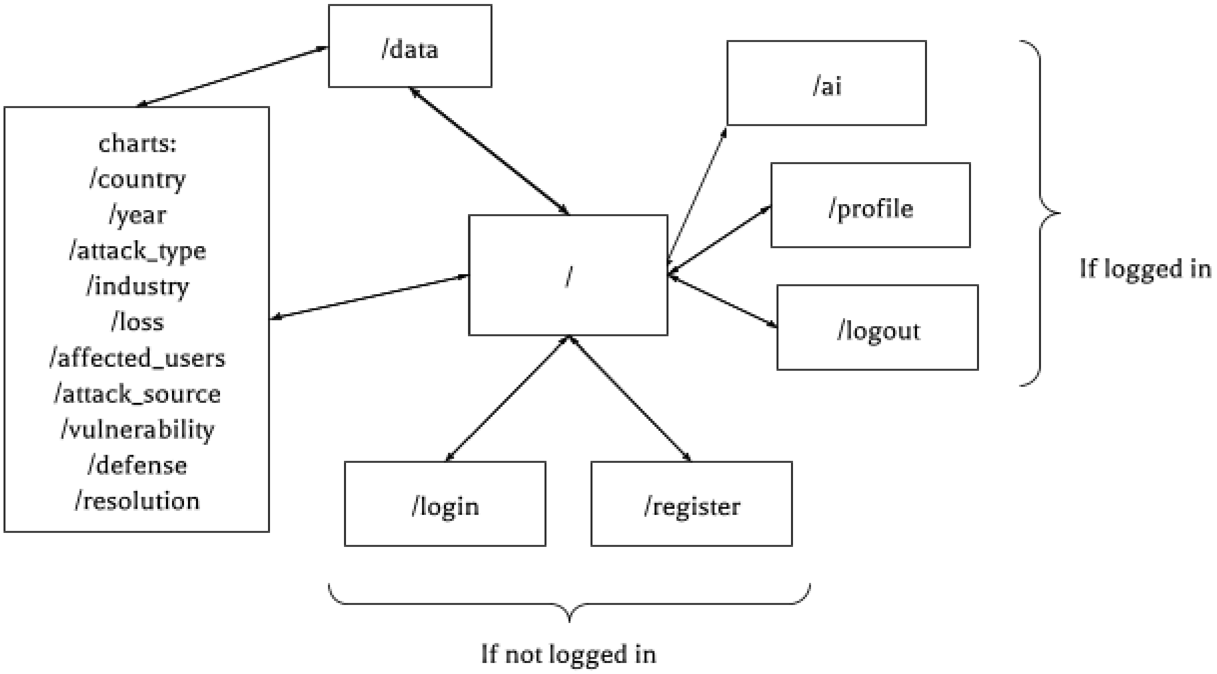
Program Components

- Login/register/logout:
    - Users do not need to register or log in to use the site, but have the option to make an account if they wish
        - If users are logged in, they can access our AI Cyberattack Predictor Model
    - Users stay logged in until they log out (Flask session)
    - Account information stored in user database
- Profile page:
    - Users are presented with a profile icon and and greeted by their username
- Home page:
    - Greeting for the user, project overview, about the database we used, basic information about how to use the site, etc.
- Data page:
    - Table with all the data loaded onto the page
    - Has a dropdown menu so users can sort by different categories (e.g., year, increasing response time, etc.)
    - Search bar allows users to filter data for specific results
- Charts pages (one for each column of the dataset):
    - Displays interactive charts that showcase different aspects of the entire dataset
- Artificial intelligence page:
    - “Predict” the details of a cyberattack based on our dataset
    - Using PyTorch, we can train the model on the cybersecurity data of the past decade so that it can make accurate predictions; for example, given the year, country, attack type, and targeted industry of a cyberattack, model can predict resolution time.
    - Display results of the AI model on backtesting on the page.
- Python/Flask components:
    - Main python file will run the entire python-flask interaction
    - Other python files will manage different pages/a python file for the AI

Component Map

```
Frontend (HTML/CSS/Bootstrap/JS)
    |
    +-- ApexCharts
    |       - Tables and charts to display data
    +-- Search and sort functionality
    v
Backend (Flask) —> login/register/logout
    |
    +-- SQL Database
    |       - User table
    |       - Cybersecurity threats table
    |
    +-- Functions for handling data
    +-- PyTorch
```

Site Map



- Navbar at the top of each page will allow for easy navigation between pages
- Home page and login/register or profile and logout pages are accessible from any page

Database Organization **(Column names with grey background, entries without)**

- Relational Database (SQLite3)

User database:

username	password
user1	*********
user2	************
...	...

- Usernames must be unique

Cybersecurity Threats database:

Country	Year	Attack Type	Targeted Industry	Financial Loss (in million $)	Number of Affected Users	Attack source	Security Vulnerability Type	Defense Mechanism Used	Incident Response Time (in hours)
China	2019	Phishing	Education	80.5377	3169	Hacker Group	Unpatched Software	VPN	63
UK	2024	Ransomware	Telecommunications	41.44	6593	Nation-state	Social Engineering	AI-based Detection	7
...	...	...	...	...	...	...	...	...	...

APIs

- We do not plan on using any APIs.

Front-End Framework: Bootstrap

Why

Bootstrap is versatile and easy to use and comes with many predefined elements. Its designs automatically adjust to different screen sizes and are highly customizable, and do not depend on CSS as much as the other FEFs do.

How

We plan to use the following Bootstrap elements:

- Top navbar to navigate the site
    - Dropdown menu for charts
- Search bar, buttons, etc. (for basic website functionality)

Data Visualization Library: ApexCharts

Why

ApexCharts allows for the creation of many different kinds of interactive charts. It has detailed documentation and allows sufficient customization without feeling overwhelming.

Task Assignments

Anastasia: JavaScript and backend development

- Generate interactive charts using visualization libraries
- Create and write functions to utilize SQLite3 database

Christopher: Flask and artificial intelligence

- Use PyTorch to train machine learning model and predict future cyber attacks
- Create home page and Flask framework

Dua: Frontend development

- Design HTML and incorporate HTML/CSS layout
- Create dataset page

Suhana: PM and data parsing

- Implement searching and sorting functionality