PIKacmT STUDENT CHAPTER

Pasckathon 2020

Theme :- Cyber Security

Team Leader Name :- Ankita Vaid **Team Name :-** SHANAYA

Problem Statement :- Effective use of Cyber Space to prevent violence against and trafficking of women and children.

College Name :- College of Engineering Pune

(This problem being tried through this solution is huge, so requires support of the Government, and hence we have assumed the Ministry of Women and Child Development (MWCD) as the authoritative and administrative organization)

What is the Problem?

-> **Violence** against **children and women** on **Social Network** takes place in these ways:

i. Posting of ***offensive images/videos*** that relate to various forms of harassment

ii. Actions deliberately undertaken for **exploitation** of children by building emotional connection

-> **Cyber Trafficking** can be broken down into two parts - **Abduction** and **Sale**. Many abduction cases and almost all sale cases are facilitated by ***the internet*** through ***fake advertisements*** (for luring victims), ***chatrooms*** etc.

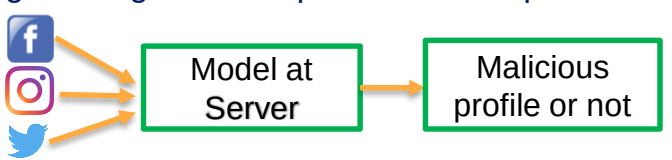
Idea/Prototype/Solution

The **Women & Child Safety Platform** is a Web Portal built to **detect all forms of violence and trafficking** against women and children by detecting fake job advertisements for waitress, model, etc., detecting chat rooms used for trafficking/ selling of children. The platform's server will continuously perform **web crawling** (visiting multiple websites) and **web scraping** (extracting data) on **Social media** and other websites to detect malicious content. It will then **automatically allot an investigator** and notify the government administrator. The platform has a built-in solution for catching hold of the trafficker (using **multilingual Chatbot Shanaya**, which **will pose as a potential victim**) and make use of data provided by the web scraper. The platform also has the facility to **allow citizens to report** such incidences of cyber trafficking to the authorities.

(On server Side) Web crawling and web scraping :

Social Media (Facebook, Twitter, Instagram etc): A crawler and scraper will be **continuously running** on social media websites to **detect suspicious profiles** (using the trained model at the server) based on:

i. The text & content in comments, postings, feeds.
ii. Images & Videos showing violence (by *image processing*).
iii. Child grooming behavior patterns of the profile.



Advertisement Websites: Crawling on job-posting websites and scraping all jobs posted. **Reverse Imaging and Data Mining Techniques** are used to extract the following data from the jobs:

a. Job title/description
b. Compensation
c. Email/phone number of the **job-provider** (suspect as trafficker)

A trained ML model at the server is used to predict if the Job title/Description is suspicious. This **extracted data is stored** in structured format **in the database,** which will be later **used by Chatbot**.

Any malicious activity/job-posting will be shown as an alert on the Investigator's Dashboard.

Multilingual Chatbot “Shanaya”

For Prevention of WOMEN Trafficking

Chatbot Shanaya will **contact job-provider (details received from scrapping the advertisement)** and pose as a victim. Shanaya uses **NLP** and a model based on **cosine similarity** to generate a suitable reply and **interact with the job providers**. The replies of the job-provider are monitored by server which analyses the replies and classifies the job-provider as suspicious or not. **If found suspicious**, appropriate **alerts** and associated information is **displayed on the investigator dashboard**.

Shanaya: Hi! Mein Shanaya hun. I got to know about your job from XYZ website, and am interested to apply for the same.

Job-Provider: Hey! What’s your ASL, body dimensions. How far will you go for money?

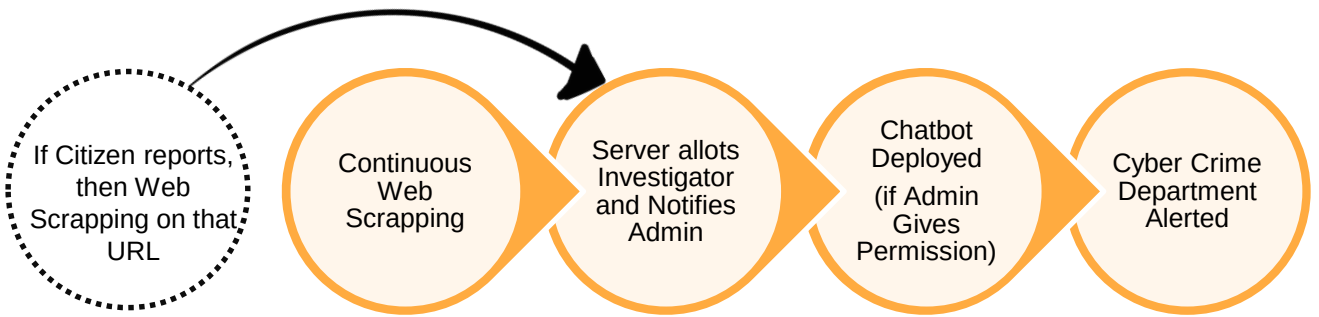
Flagged as MALICIOUS

For Prevention of CHILD Trafficking

Chat rooms: The **chatbot poses as a 13-year-old child** and engages in conversations in supposedly malicious chatrooms(E.g. having titles with keywords like “**secks**”(sex)). Consequently, if the conversation is found to be malicious by the trained model, such **chat rooms will be reported to the cyber crime department** of India who can block them and not allow creation of such chat rooms with similar keywords.

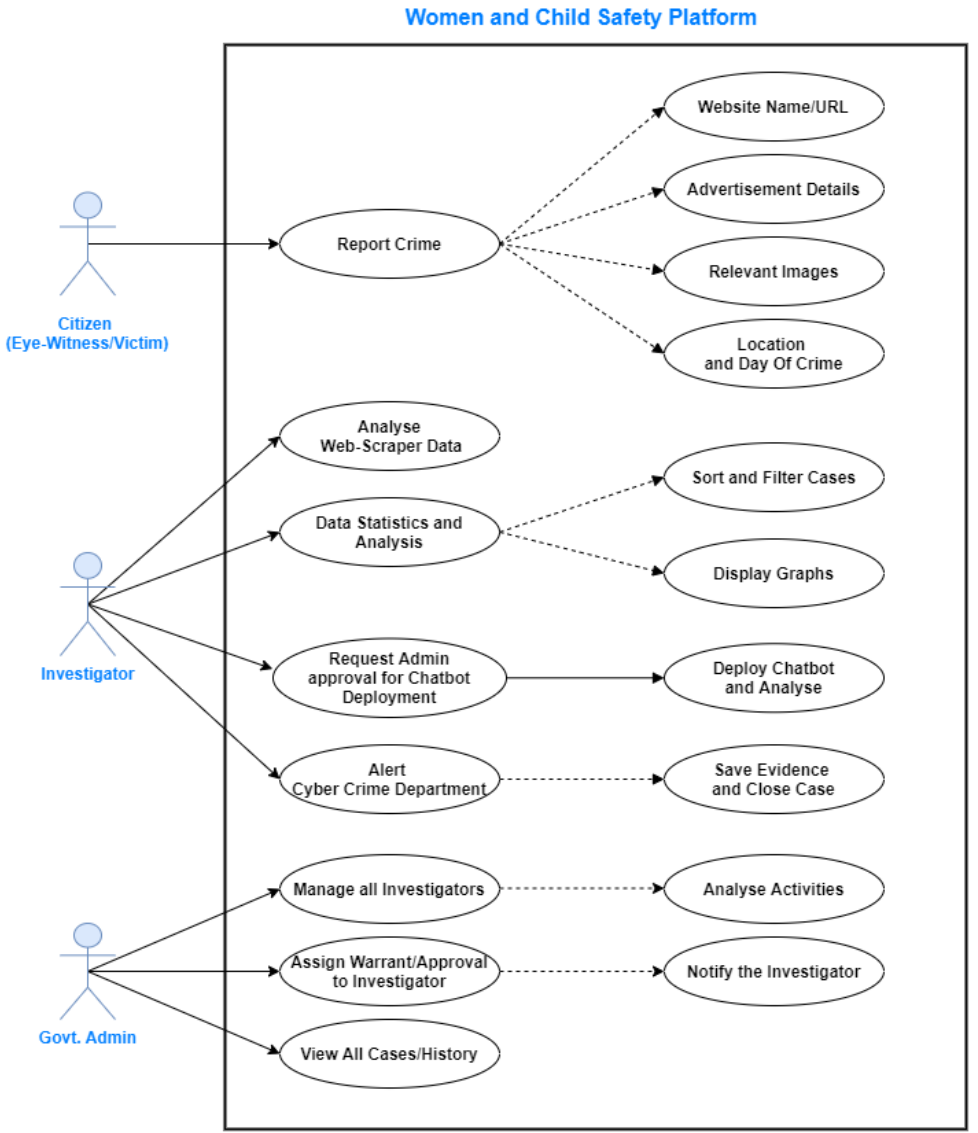
The server **will automatically allocate** the cases to the investigator & display on the dashboards.

Timeline:



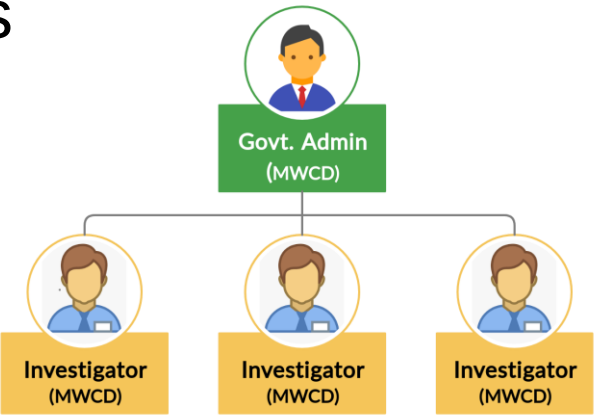
Technology Stack

- **Front End**
  - HTML5,CSS3
  - Bootstrap
  - JavaScript
- **Backend**
  - PHP
  - Python
- **Database**
  - MongoDB
- **Libraries**
  - Scrappy
  - Selenium, BS4
  - NLTK
  - Twilio/Plivo API
  - OpenCV
  - Animate.css



Use Cases

The Platform is modularized and provides access according to the following **3 views** -



MWCD = Ministry of Women & Child Development

Citizen view:

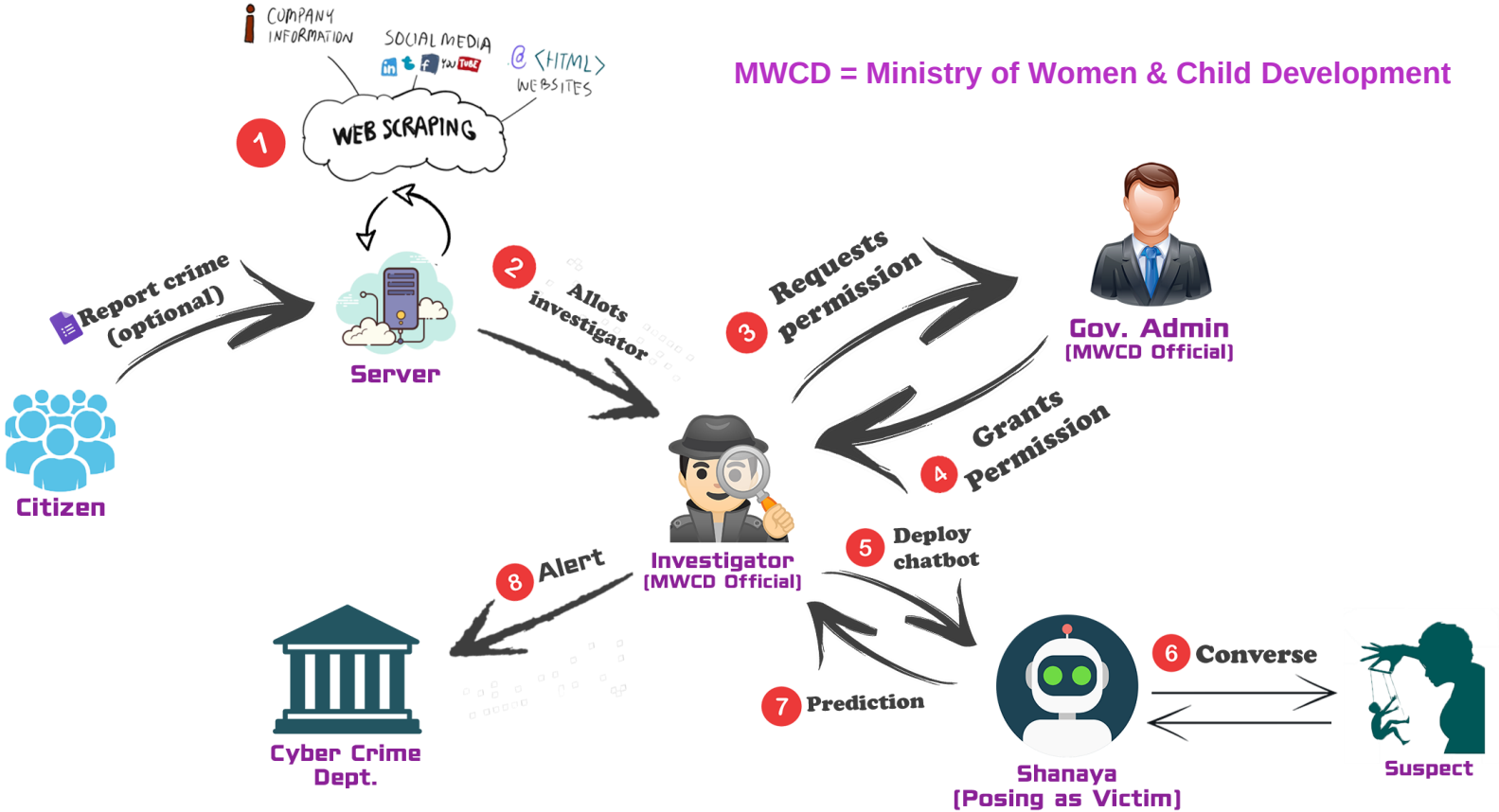
1. Citizen can ***report crime*** as an eye-witness or a victim.
2. They can **choose** to remain **anonymous.**

Investigator Login: (Lower ranked Ministry officials)

1. Analyse **user-form data** and **scrapper output**.
2. In case of malicious activity - request (Admin) approval to deploy chatbot.
3. If approval granted, deploy chatbot.
4. Analyse outcome of chat, alert Cyber Crime Dept. if found malicious.
5. Dashboard has options to **view, analyse and sort cases**.
6. Generate and export **case reports**, graphs and other statistics.

Government Administrator Login:

1. Admin access is the **highest ranked authorization at the ministry**.
2. Complete access to all the case files and **generate statistical reports.**
3. Overlooks investigators and **monitors their activities**.
4. Grant **permissions** to investigator to deploy chatbot “Shanaya” and conduct further investigation.



MWCD = Ministry of Women & Child Development

FLOW DIAGRAM

Dependencies / Show Stoppers

1. Data from previous records, which contains details of criminals(name, age, contact number), area in which he committed crime, his modus operandi(with exact words he/she used for victimizing women/children).

2. Data Set for training Chatbot Shanaya and the prediction model is not available, so we have **created our own dataset.**