SMART INDIA HACKATHON 2023

Basic Details of the Team and Problem Statement

Ministry/Organization Name/Student Innovation:
National Technical Research Organisation (NTRO)

PS Code: SIH1677

Problem Statement Title: Developing a tool to provide for real time feeds of cyber incident pertaining to Indian Cyber Space.

Team Name: HackSmiths

Team Leader Name: Harsh Kumar

Institute Name: Netaji Subhas University of Technology

Idea/Approach Details

Problem Overview

The National Critical Information Infrastructure Protection Centre (NCIIPC) shares details of cyber incidents with relevant stakeholders to protect India's Critical Information Infrastructure (CII). These real-time updates on cyber activities allow stakeholders to take relevant measures to mitigate risks.

Our Approach

- Use Machine Learning to identify platforms that share cyber incident data.
- Develop a framework to collect cyber incident data from identified platforms.
- Create a well-structured database of cyber incidents.
- Generate insights and visual representations of cyber incidents categorized by sectors, Advanced Persistent Threats (APTs), and strategic issues.

Objective

To develop a real-time feed tool for cyber incidents in Indian cyberspace. This tool will enhance threat assessment capabilities by collecting and analyzing cyber incidents reported across the web, including forums, social media, paste sites, and other platforms.

Problem Overview

Description of our Solution

- **Data Collection & Scraping:** Collecting data on cyber incidents from various online platforms, such as social media, forums etc. This will include gathering both structured and unstructured data.

- **Data Warehouse and Structuring:** Building a data warehouse to store both historical and real-time data. The data will be structured in a suitable format for analysis and stored in a knowledge graph to represent relationships and entities.

- **Multi-Agent System:** Developing a multi-agent system to analyze data in real-time. The system will consist of:
  a. **Live Info Accumulator:** Combining of scraping tools to gather real-time data on cyber incidents.
  b. **Data Processing Module:** Utilising ML and NLP to analyze the patterns, identify potential threats, and derive insights on socio-economic impacts.
  c. **Interface Module:** A full-stack web interface for authorities to interact with the system, view real-time updates, and generate reports.

- **Two-Level Analysis:**
  a. **Level 1:** Classify the scraped data to determine if it is harmful or linked to larger events, then store it.
  b. **Level 2:** Analyze data to detect potential larger events or chronological patterns that could indicate escalating threats.

Idea/Approach Details

Problem Overview

The National Critical Information Infrastructure Protection Centre (NCIIPC) shares details of cyber incidents with relevant stakeholders to protect India's Critical Information Infrastructure (CII). These real-time updates on cyber activities allow stakeholders to take relevant measures to mitigate risks.

Our Approach

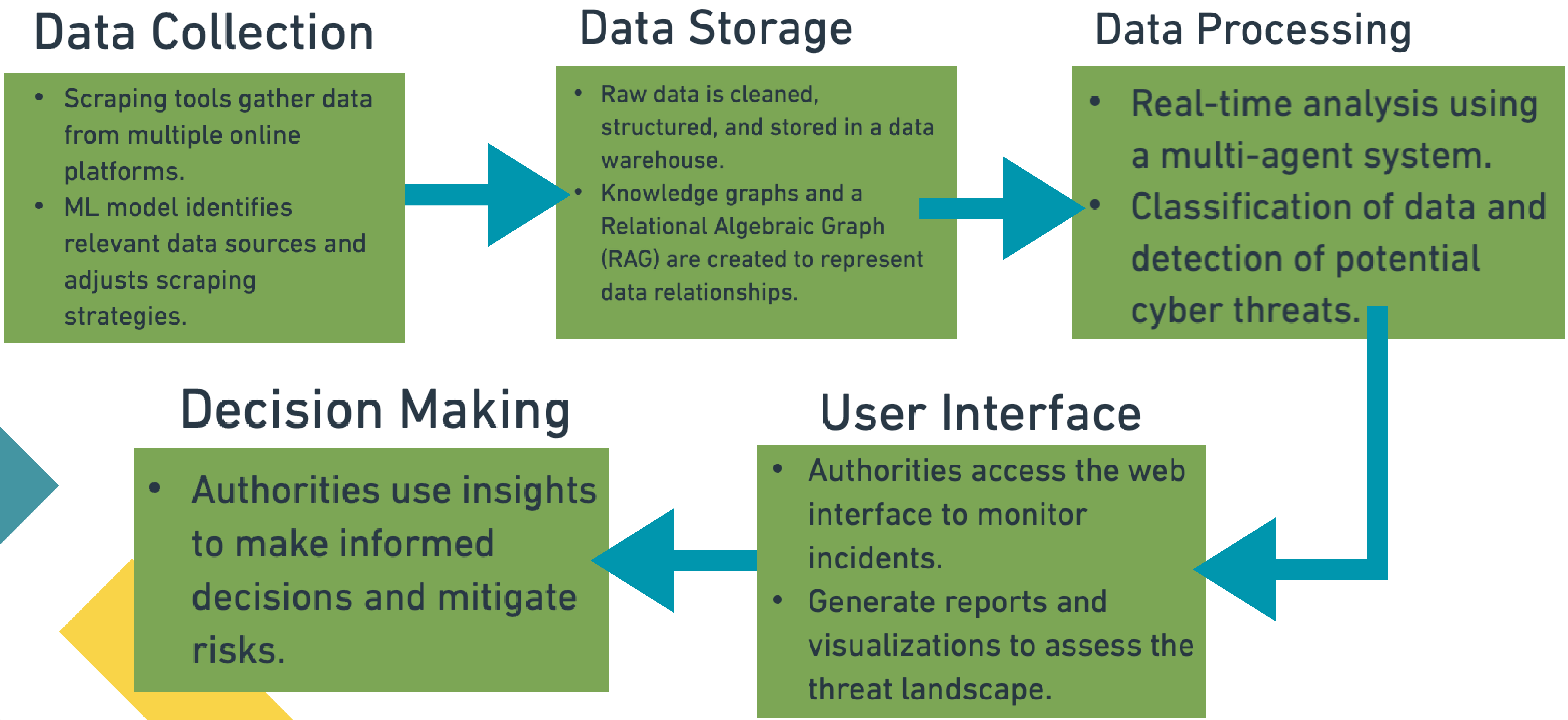
- Use Machine Learning to identify platforms that share cyber incident data.
- Develop a framework to collect cyber incident data from identified platforms.
- Create a well-structured database of cyber incidents.
- Generate insights and visual representations of cyber incidents categorized by sectors, Advanced Persistent Threats (APTs), and strategic issues.

Objective

To develop a real-time feed tool for cyber incidents in Indian cyberspace. This tool will enhance threat assessment capabilities by collecting and analyzing cyber incidents reported across the web, including forums, social media, paste sites, and other platforms.

Idea/Approach Details

Data flow chart of the platform



Data Collection

- Scraping tools gather data from multiple online platforms.
- ML model identifies relevant data sources and adjusts scraping strategies.

Data Storage

- Raw data is cleaned, structured, and stored in a data warehouse.
- Knowledge graphs and a Relational Algebraic Graph (RAG) are created to represent data relationships.

Data Processing

- Real-time analysis using a multi-agent system.
- Classification of data and detection of potential cyber threats.

User Interface

- Authorities access the web interface to monitor incidents.
- Generate reports and visualizations to assess the threat landscape.

Decision Making

- Authorities use insights to make informed decisions and mitigate risks.

Expected Tech Stack

- **Data Colection & Scraping:** Python, Scrapy, Selenium
- **Machine Learning:** TensorFlow, PyTorch, Scikit-Learn
- **NLP:** NLTK, Hugging Face Transformers
- **Data Storage:** PostgreSQL, MongoDB
- **Backend:** Flask, FastAPI
- **Multi-Agent Framework:** CREWAI(Custom Framework)
- **Deployment:** AWS EC2, Docker
- **Frontend:** React, Nex.js, Tailwind CSS
- **Additional Tools:** LangChain, OpenAI LLM, Kafka

Miscellaneous

Business Viabilty

- **Market Demand:** Increasing cyber threats to critical infrastructure make this tool essential for national security.
- **Cost-Effectiveness:** Open-source tools and in-house data processing reduce costs, making the solution affordable.
- **Scalability:** The system is designed to handle large-scale data and can be expanded to cover new threat vectors.
- **Competitive Advantage:** Real-time insights and a comprehensive database provide a unique advantage in threat assessment and mitigation.
- **Revenue Model:** Potential monetization through government contracts, cybersecurity firms, and international collaborations.

Team Member Details

Team Leader Name: Harsh Kumar

Branch: BTech Stream: CSAI Year: IV

Team Member 1 Name: Avneet Singh Bedi

Branch: BTech Stream: CSAI Year: IV

Team Member 2 Name: Balvinder Singh

Branch: BTech Stream: CSAI Year: IV

Team Member 3 Name: Laksshay Sehrawat

Branch: BTech Stream: CSAI Year: IV

Team Member 4 Name: Raghav Mangla

Branch: BTech Stream: CSAI Year: IV

Team Member 5 Name: Sneha Gupta

Branch: BTech Stream: CSAI Year: III