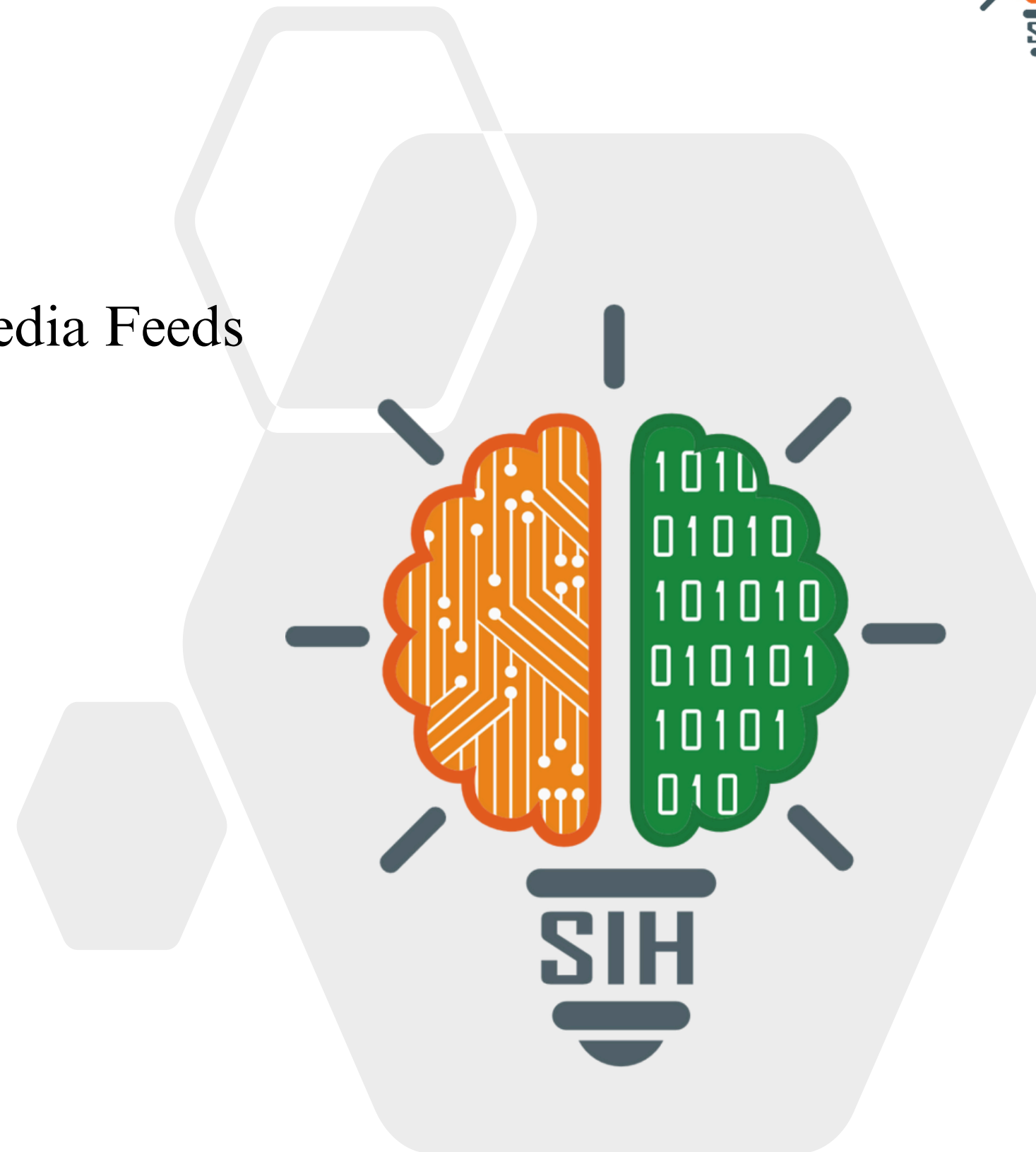
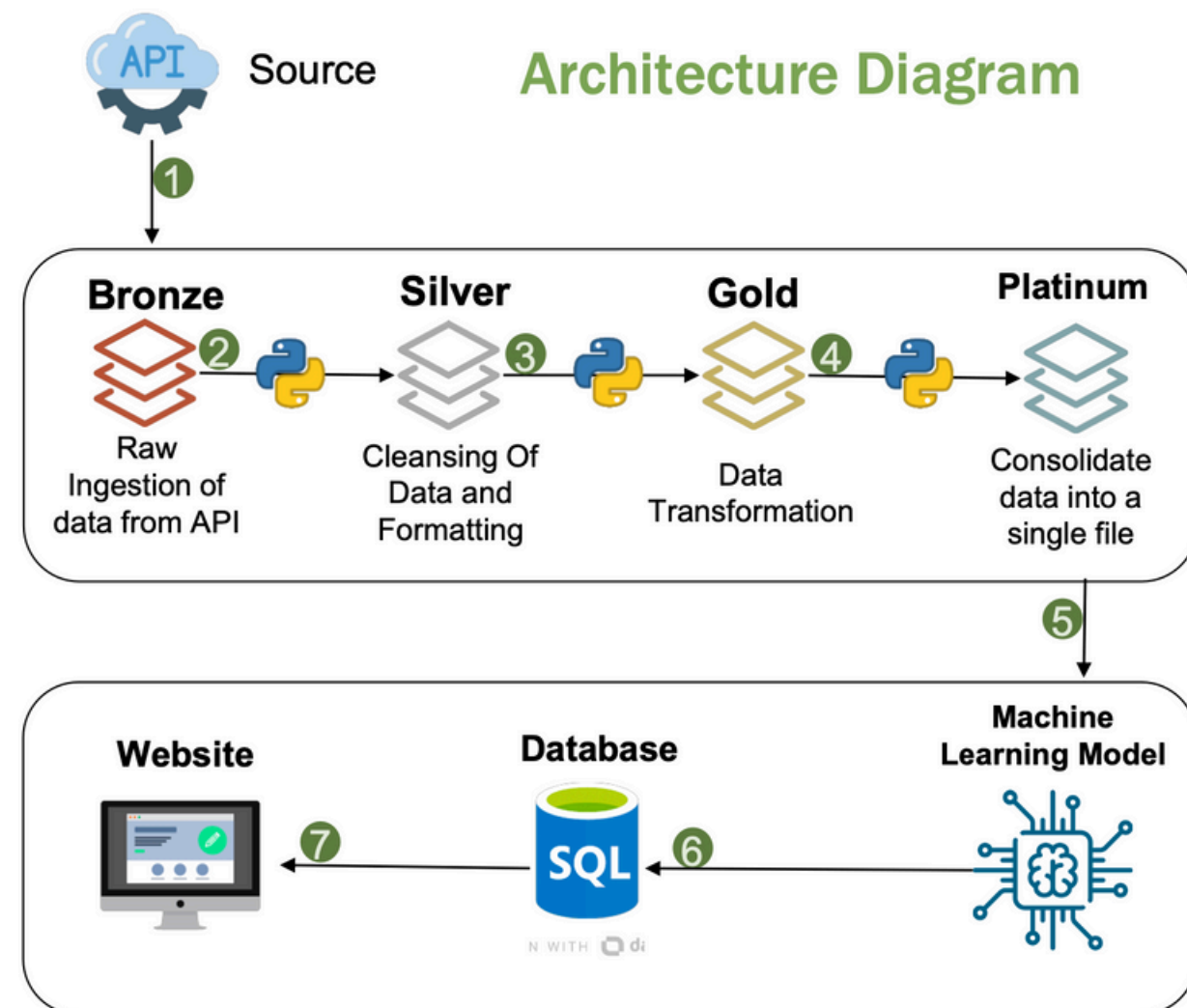


- **Problem Statement ID** - 1743
- **Problem Statement Title** - Parsing of Social Media Feeds
- **Theme** - Miscellaneous
- **PS Category** - Software
- **Team ID** - 21
- **Team Name** - Matrix\_Manipulators



## Automated User Data Extraction Using Web Scraping



**Programming Languages:** Python, ReactJS, Vanilla JavaScript, AJAX

**Software/Frameworks:** BeautifulSoup4, Selenium, Tweepy, Flask, Pandas, Jupyter Notebook, VS Code

### Detailed Explanation of the Proposed Solution:

The project focuses on parsing social media feeds for forensic investigation across multiple platforms, starting with **Twitter (prototype)** and later expanding to others.

The solution involves using a combination of Python libraries such as **ntscraper** for **Nitter** (a Twitter interface), **BeautifulSoup**, and **Selenium** to extract and analyze posts and user data.

### Innovation and Uniqueness of the Solution:

**Comprehensive Platform Support:** This solution is designed to scrape multiple platforms like Twitter, Facebook, Instagram, Telegram, and WhatsApp making it versatile and widely applicable in forensic investigations.

The tool will allow **real-time data scraping** and provide forensic documentation through screenshots and data export, enabling investigators to capture critical information as it happens.

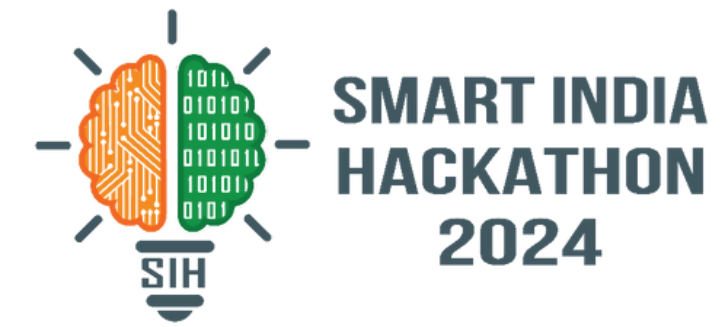
The final tool can be packaged into a **user-friendly web-based interface**, where investigators can input **usernames, hashtags, or keywords** to automatically gather the data they need, making it an innovative and accessible tool.

[Link to Github](#)

## Matrix Manipulators

# TECHNICAL APPROACH

## Methodology and Process for Implementation



### Twitter Parser

Enter User Details:

Submit

Output:

#### User Details

Image: [https://pbs.twimg.com/profile\\_images/1591558315254890500/ETB8u4NI\\_400x400.jpg](https://pbs.twimg.com/profile_images/1591558315254890500/ETB8u4NI_400x400.jpg)

Name: Jeff Bezos

Username: @JeffBezos

Id: 15506669

Bio: Amazon, Blue Origin, Washington Post, Bezos Earth Fund, Bezos Academy.

Location:

Website: <https://hoo.be/jeffbezos>

Joined: 10:38 PM - 20 Jul 2008

Stats:

Tweets: 399


Following: 384

Followers: 6393691

Likes: 178

Media: 0

Image:



Raw JSON Data:

```
{  "userdetails": {    "image": "https://pbs.twimg.com/profile_images/1591558315254890500/ETB8u4NI_400x400.jpg",    "name": "Jeff Bezos",    "username": "@JeffBezos",    "id": "15506669",    "bio": "Amazon, Blue Origin, Washington Post, Bezos Earth Fund, Bezos Academy.",    "location": "",    "website": "https://hoo.be/jeffbezos",    "joined": "10:38 PM - 20 Jul 2008",    "stats": {      "tweets": 399,      "following": 384,      "followers": 6393691,      "likes": 178,      "media": 0    }  }}
```

User Name:

Number of Tweets:

Panchanama Formatter


Generate Report

Check Stats

Download Report

User Profile (profile picture, name, username, bio, date of joining, website link)  
Statistics: Total tweets, following, followers, likes and media,  
No. of tweets for specified timeline

### Jeff Bezos - Profile Overview



Name: Jeff Bezos  
Username: @JeffBezos  
Bio: Amazon, Blue Origin, Washington Post, Bezos Earth Fund, Bezos Academy.  
Joined: 10:38 PM - 20 Jul 2008  
Website: <https://hoo.be/jeffbezos>

Statistics:

Total Tweets: 399

Following: 385

Followers: 6393529

Likes: 178

Media: 0

Recent Tweets:

Date: Jul 14, 2024 - 2:38 AM UTC  
Tweet: Our former President showed tremendous grace and courage under literal fire tonight. So thankful for his safety and so sad for the victims and their families.  
Link: <https://twitter.com/JeffBezos/status/1812315786507088370#m>  
Likes: 205052 | Comments: 8976

Date: Oct 27, 2023 - 2:00 PM UTC  
Tweet: Impressive visit to the @blueorigin Huntsville Engine Production Facility! @NASA is proud to partner with Blue Origin, especially on the Blue Moon human landing system, which will help ensure a steady cadence of astronauts on the Moon to live and work before we venture to Mars.  
Link: <https://twitter.com/SenBillNelson/status/1717904201617908137#m>  
Likes: 4809 | Comments: 2989

Date: Jun 8, 2023 - 1:05 AM UTC  
Tweet: Nominal run!  
Link: <https://twitter.com/torybruno/status/1666612451440222209#m>  
Likes: 4137 | Comments: 1282

Date: May 19, 2023 - 3:06 PM UTC  
Tweet: Honored to be on this journey with @NASA to land astronauts on the Moon — this time to stay. Together, we'll be solving the boil-off problem and making LOX-LH2 a storable propellant combination, pushing forward the state of the art for all deep space missions. #Artemis #MoonToMars <https://www.blueorigin.com/news/nasa-selects-blue-origin-for-mission-to-moon>  
Link: <https://twitter.com/JeffBezos/status/1659576309432090625#m>  
Likes: 13861 | Comments: 5766

User Name:

Number of Tweets:

Submit

Fetches Tweets:

Tweet Text: Our former President showed tremendous grace and courage under literal fire tonight. So thankful for his safety and so sad for the victims and their families.  
Date: Jul 14, 2024 - 2:38 AM UTC  
Likes: 205041  
Comments: 8981  
[View Tweet](#)

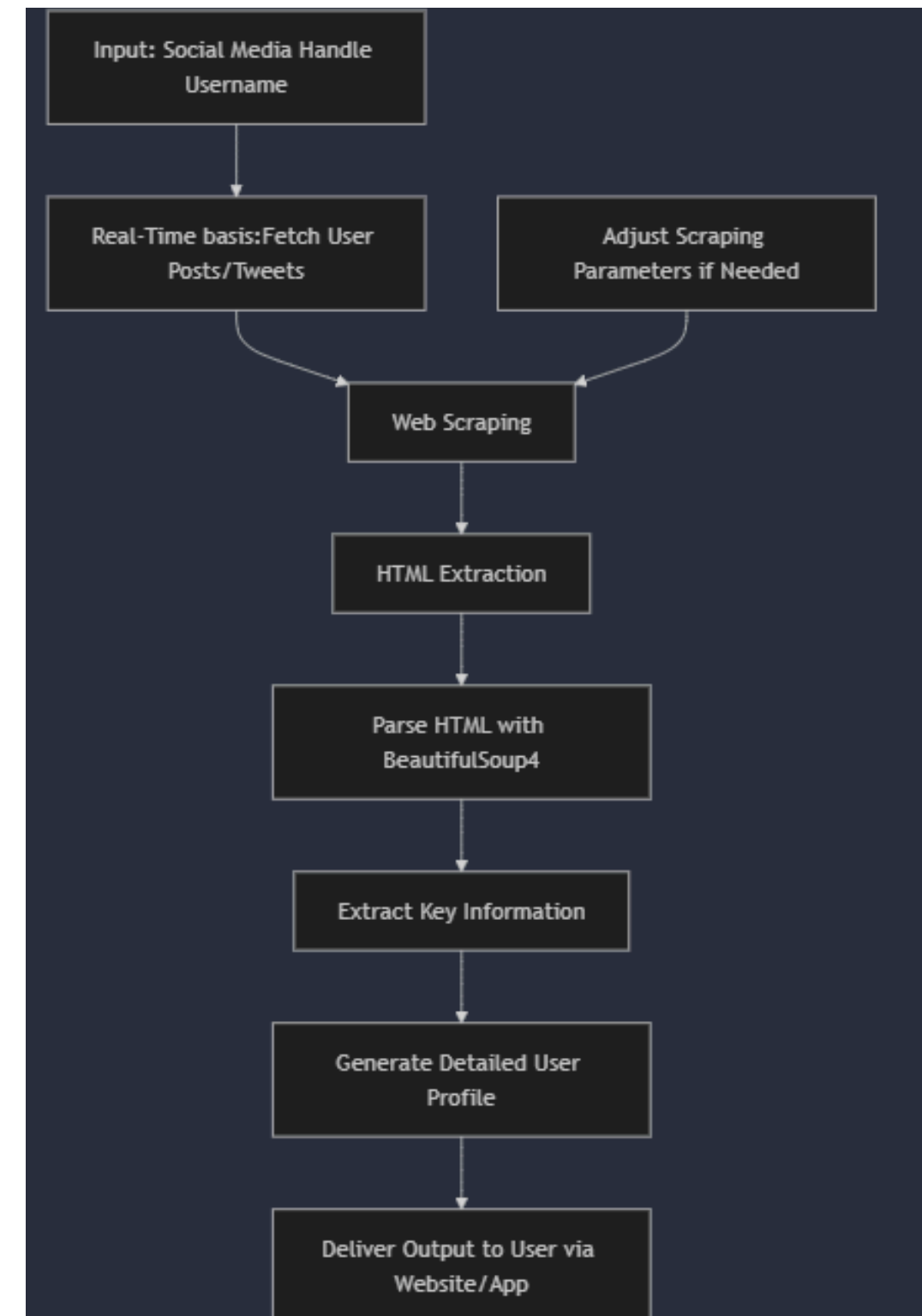
Tweet Text: Impressive visit to the @blueorigin Huntsville Engine Production Facility! @NASA is proud to partner with Blue Origin, especially on the Blue Moon human landing system, which will help ensure a steady cadence of astronauts on the Moon to live and work before we venture to Mars.  
Date: Oct 27, 2023 - 2:00 PM UTC  
Likes: 4813  
Comments: 2989  
[View Tweet](#)

Tweet Text: Nominal run!  
Date: Jun 8, 2023 - 1:05 AM UTC  
Likes: 4136  
Comments: 1282  
[View Tweet](#)

Tweet Data:

```
{  "twitter_link": {    "0": "https://twitter.com/JeffBezos/status/1812315786507088370#m",    "1": "https://twitter.com/SenBillNelson/status/1717904201617908137#m",    "2": "https://twitter.com/torybruno/status/1666612451440222209#m"  },  "text": {    "0": "Our former President showed tremendous grace and courage under literal fire tonight. So thankful for his safety and so sad for the victims and their families.",    "1": "Impressive visit to the @blueorigin Huntsville Engine Production Facility! @NASA is proud to partner with Blue Origin, especially on the Blue Moon human landing system, which will help ensure a steady cadence of astronauts on the Moon to live and work before we venture to Mars.",    "2": "Nominal run!"  },  "date": {    "0": "Jul 14, 2024 - 2:38 AM UTC",    "1": "Oct 27, 2023 - 2:00 PM UTC",    "2": "Jun 8, 2023 - 1:05 AM UTC"  },  "likes": {    "0": 205041,    "1": 4813,    "2": 4136  },  "comments": {    "0": 8981,    "1": 2989,    "2": 1282  }}
```

[Copy Tweets](#)





## Feasibility Analysis

- **Technical Feasibility:** Leveraging Python's extensive scraping and automation libraries ensures effective data retrieval from social media platforms.
- **Operational Feasibility:** Automating the creation of screenshots and reports reduces human error and improves efficiency in forensic investigations.
- **Economic Feasibility:** Minimal infrastructure costs, given the reliance on open-source tools and technologies, make this solution cost-effective.

## Potential Challenges and Solutions

**Platform Restrictions:** Many social media platforms (Facebook, Instagram) have strict anti-scraping policies or may block bots, potentially limiting access to data. Used **API's for structured access** to data.

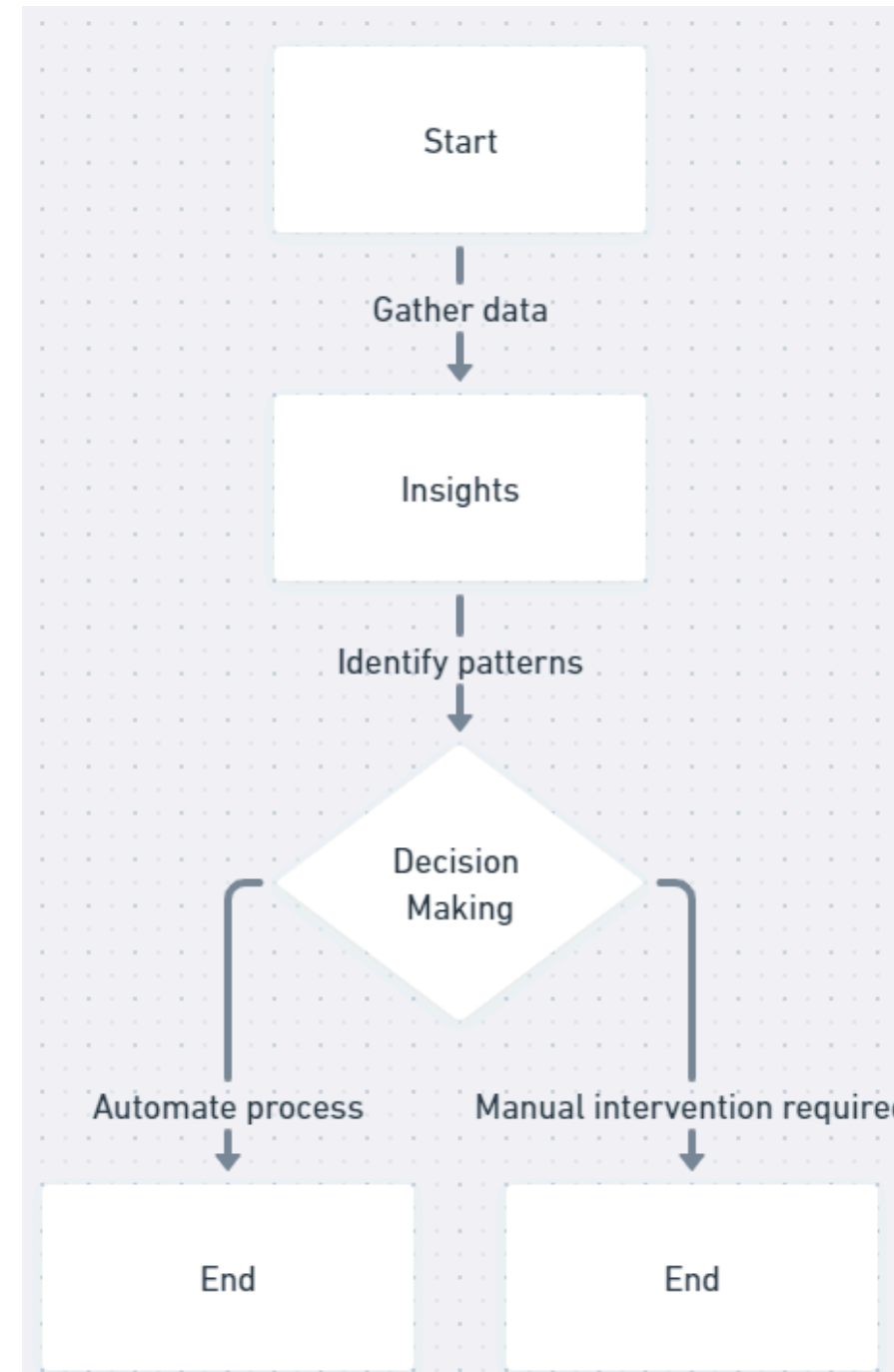
**Captcha mechanisms or two-factor authentication (2FA)** can disrupt the scraping process, making it harder to automate data collection. Integrated AI/ML techniques for CAPTCHA solving and set up pre-authorized credentials on dummy accounts for 2FA.

**Dynamic Content Loading:** Social media platforms often use JavaScript for dynamic content loading (infinite scroll), making it difficult to extract all posts using basic scraping techniques. Used selenium to handle JavaScript-rendered content.

**Legal and Ethical Considerations:** Scraping social media data, especially for forensic purposes, may involve legal challenges related to privacy and data protection laws. Ensured all data retrieval adheres to privacy policies and legal guidelines.

## Potential Impact on Target Audience

- **Users and Analysts:** The solution provides valuable insights into social media activities, enabling users and analysts to gather comprehensive information about specific individuals. This can be beneficial for personal research, market analysis, and understanding social media trends.
- **Informed Decisions:** Access to detailed and accurate social media data supports better decision-making. Whether for research, marketing, or personal use, having reliable data helps make informed choices and strategies.
- **Automation:** The automated nature of the bot reduces the time and effort required to manually gather and process social media data. This leads to increased efficiency and productivity for users and organizations.



## Benefits of the Solution

- **Automating the parsing of social media feeds** allows investigators to focus on data analysis rather than manual data collection, saving time and effort.
- The solution works across **multiple social media platforms** (Twitter, Facebook, Instagram, Telegram, WhatsApp, etc.), and its **dual versions for Android and Windows** provide flexibility in handling mobile-based and desktop-based platforms.
- Investigators can **extract specific details** such as posts, messages, follower lists, etc., and receive outputs in various formats (e.g., screenshots, PDF reports) tailored to their needs.
- This tool significantly **reduces delays** caused by manual methods, expediting the overall investigation process.

- **"Tweepy Documentation"** - Documentation for Tweepy, a Python library for accessing the Twitter API. <https://docs.tweepy.org/en/stable/>
- **Fast API:** <https://fastapi.tiangolo.com>
- **Beautiful Soup:** <https://beautiful-soup-4.readthedocs.io/en/latest/>
- **Selenium:** <https://www.selenium.dev>