NRIIT HACKATHON 4.0

- Team ID : 21
- Problem Statement Title : Instagram Spam Detection System for Enanced User Experience
- Theme : Full Stack Development

Introduction

In today's social media environment, spam content is a significant challenge, especially on platforms like Instagram. Spam comments, often promoting phishing attempts or fake offers, negatively impact user engagement and security. Our project, Instagram Spam Detection System, aims to solve this issue by automating the detection and categorization of spam comments under Instagram posts and reels. The system uses AI-based techniques to detect spam in real-time, enhancing the overall user experience on the platform.

Problem Statement

Instagram users often encounter spam comments under their posts and reels. These comments can include:

- Fake Promotions: Irrelevant ads or offers that distract users.
- Phishing Scams: Attempts to steal personal information.
- Irrelevant Content: Unwanted or repetitive messages that harm user engagement.

Given the large volume of posts and comments, manually detecting these spam comments is a tedious and time-consuming process. Instagram lacks an automated system to filter out spam comments effectively, which led to the development of our solution.

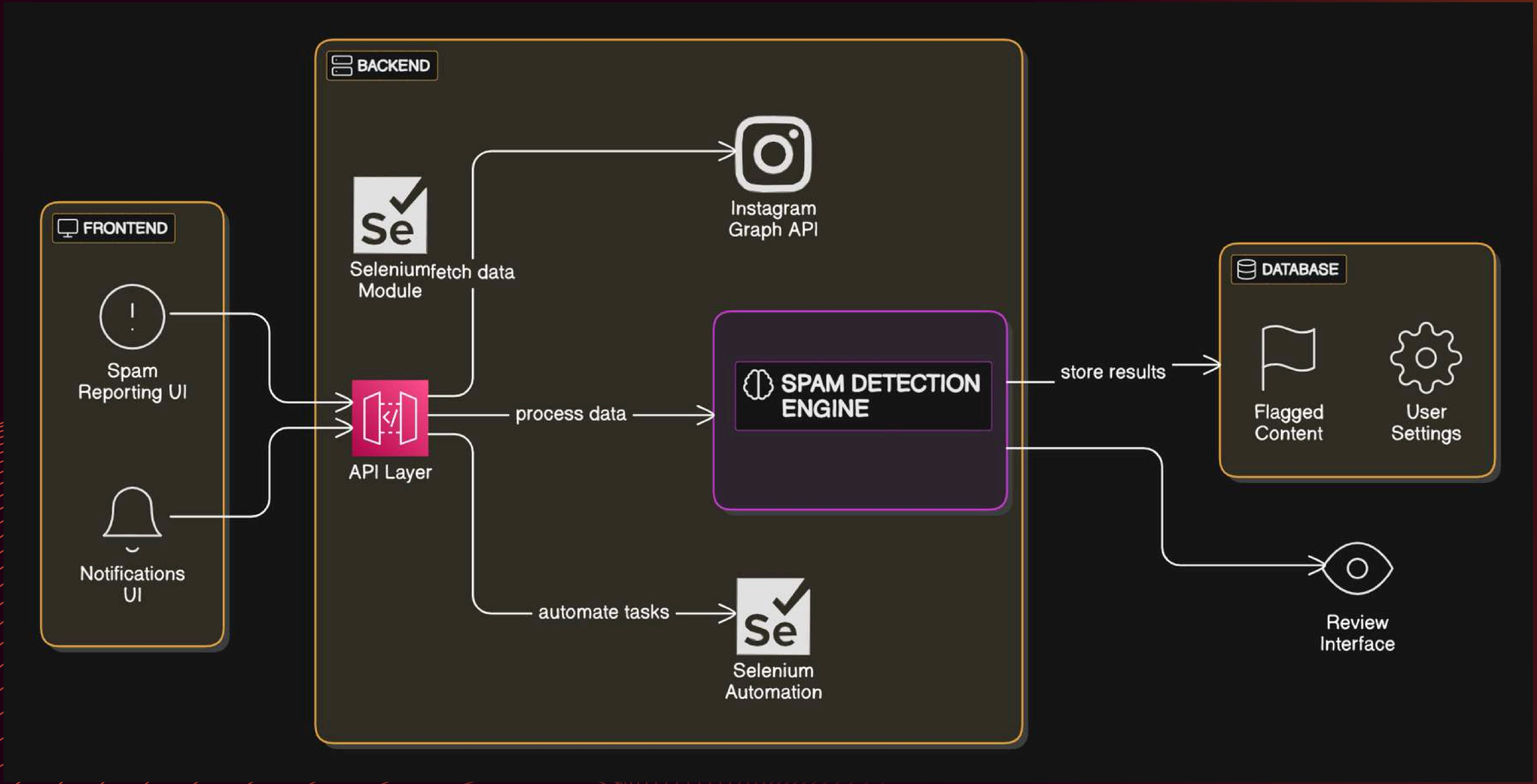
Project Objectives

The main objectives of this project are:

- AI-Powered Spam Detection: To automatically detect spam comments under Instagram posts and reels using AI techniques.
- Categorization: Classifying comments into high, medium, and low-risk categories based on their content.
- Real-Time Detection: The system detects spam comments in real-time and alerts users.
- User Control: Enabling users to review flagged comments and take action (approve or dismiss).

By automating spam detection, we aim to improve user engagement and provide a safer social media experience.

Architecture Overview



Methodology

Step 1: Setting Up the Environment

- We use Selenium WebDriver to automate interactions with Instagram. Selenium simulates user behavior like opening posts and reels, which helps fetch comments from the platform.
- We also integrate the Instagram Scraper API to retrieve comments from Instagram posts and reels programmatically. This API allows us to fetch comment data in real-time without complex scraping.

The combination of Selenium and RapidAPI enables efficient data extraction and interaction with Instagram.

Step 2: Spam Detection

Spam comments are detected using a keyword-based approach. The system scans the text of each comment and checks for keywords that indicate spam:

- High-Risk Keywords: Words like "scam", "fake", and "offer", which are indicative of harmful or deceptive content.
- Medium-Risk Keywords: Words like "free", "deal", and "discount", which suggest suspicious content.
- Low-Risk Comments: Normal comments that do not contain spam-related keywords.

Step 3: Categorization

Once a comment is flagged as potential spam, the system categorizes it into:

- High Risk: Comments containing harmful content, such as scams or phishing attempts.
- Medium Risk: Suspicious comments that might be spam but are less harmful.
- Low Risk: Normal, non-spam comments.

Methodology

The project utilizes several key technologies:

- Selenium WebDriver: For automating Instagram interactions, like opening posts and reels, and scraping comments.
- RapidAPI Instagram Scraper API: Fetches comments from Instagram posts and reels.
- Python Libraries:
  - http.client: Makes HTTP requests to retrieve data.
  - json: Handles the JSON data returned by the Instagram Scraper API.
  - time: Used for adding delays between actions to ensure smooth execution.

Results and Testing

The system was successfully tested on multiple Instagram posts and reels. It effectively categorized spam comments into high, medium, and low-risk categories. The system flagged comments in real-time and allowed users to review them.
Some of the challenges faced during testing:

- Dynamic Content: Instagram loads comments dynamically as users scroll. This required the use of Selenium to automate scrolling and load all comments.
- API Rate Limits: The Instagram Scraper API has rate limits, so careful management of API requests was necessary to avoid hitting those limits.

Conclusion and Future Work

Conclusion:

- The Instagram Spam Detection System automates the detection and categorization of spam comments under Instagram posts and reels. It categorizes comments into risk levels, enabling users to manage flagged content and improving the overall user experience.

Future Work:

- Advanced Models: Use models like BERT or GPT-3 for better spam detection.
- Cross-Platform Integration: Extend the solution to platforms like Facebook and Twitter.
- Behavioral Analysis: Detect fake accounts and bot-like behavior based on comment patterns.