INEXTURE

ScrapSync

Case Study

www.inexture.com 2024

Table of Contents

Project Overview

The Vahan Scraping App is a Python-based web application developed using the Flask framework, designed to streamline the process of extracting vehicle details from the official Indian government's Parivahan website. (https://vahan.parivahan.gov.in).

This app serves as a gateway for individuals and businesses alike, providing easy access to critical information related to vehicle registration, ownership, and company.

Through the strategic implementation of advanced web scraping techniques, the application autonomously navigates the website's pages after login, intelligently parsing HTML elements to extract information with remarkable precision and efficiency.

About Customer

Our customer, a leading data analytics firm, seeks to develop a comprehensive data scraping solution tailored for their client's needs, extracting information from the Vahan Parivahan portal of the Indian government. The solution aims to gather essential data such as vehicle owner details, vehicle type, registration numbers, insurance particulars including type and expiry, PUC (Pollution Under Control) status, and other related vehicle information.

This solution serves multiple use cases, including vehicle registration verification, insurance monitoring for compliance, pollution control adherence, and facilitating smooth vehicle ownership transfers. By leveraging this scraped data, businesses and stakeholders can conduct market research to understand consumer preferences and market trends, while also aiding in traffic management strategies, safety initiatives, and government policy formulation related to transportation and environmental conservation.

Additionally, the data can be utilized to offer customized services such as insurance quotes and vehicle maintenance reminders. However, it's imperative to ensure strict adherence to data privacy regulations and obtain explicit user consent for data collection and processing, thereby upholding ethical standards throughout the product development and implementation process.

Purpose

The purpose of the Vahan Scraping App is to simplify the retrieval of vehicle details for users who require information such as registration status, vehicle class, fuel type, emission norms, owner details, and more. By automating the process of data extraction from the Parivahan website, the application aims to save time and effort for individuals and businesses involved in various activities related to vehicles, including buying/selling, insurance, and regulatory compliance.

The users are required to input the vehicle number into the application's API which then seamlessly extracts all the relevant details from the Parivahan website after the login process is completed, eliminating the need for manual data entry and providing users with instant access to comprehensive vehicle information.

This streamlined workflow not only enhances user convenience but also enhances the efficiency and accuracy of accessing critical vehicle-related data, ultimately facilitating informed decision-making and optimizing operational workflows.

Business Challenge

The primary challenge faced by individuals and businesses in India is the lack of a centralized platform for accessing comprehensive vehicle information. The existing process of obtaining such information involves navigating through multiple government portals and often requires manual intervention, leading to inefficiencies and delays. Moreover, the Parivahan website itself may not provide an API for accessing vehicle data programmatically, necessitating the development of a scraping solution. The absence of a centralized platform for accessing comprehensive vehicle details prompts the development of the Vahan Scraping App. The challenges were as follows:

1. Manual Data Retrieval: The existing process of retrieving vehicle information involves manual entry and navigation through multiple web pages, resulting in time-consuming and error-prone data retrieval processes.

2. Limited Accessibility: Individuals and businesses face limitations in accessing vehicle details like the Parivahan portal only limit 5 vehicle details per login session.

The goal of the Vahan Scraping App project is to address these challenges by providing a user-friendly and efficient platform for accessing comprehensive vehicle information from the Indian government's Parivahan website. By automating the process of data extraction and retrieval, the application aims to streamline workflows, enhance accessibility and ultimately empowering users with timely and accurate vehicle details for informed decision-making and operational efficiency.

Our Solution

The Vahan Scraping App addresses the business challenge by offering a user-friendly API for retrieving vehicle details directly from the Parivahan website. Leveraging web scraping techniques, the application programmatically extracts relevant information from the website's HTML pages, bypassing the need for manual data entry or navigation. The Flask framework provides the backbone for the application, offering a lightweight and modular structure for building web applications in Python. With its intuitive design and efficient data retrieval capabilities, the Vahan Scraping App offers a reliable solution for accessing up-to-date vehicle information.

Key Functionalities

1. Robust Data Retrieval Engine:

Powered by Python and integrated with web scraping library such as Selenium, the application boasts a robust data retrieval engine capable of extracting comprehensive vehicle details from the Parivahan website. Through automated processes, users can seamlessly retrieve information such as registration status, vehicle class, owner details, and more with minimal manual intervention.

2. User-Friendly:

The Vahan Scraping App features a user-friendly API designed for ease of navigation and intuitive operation. Users can easily input vehicle number, initiate data retrieval processes, and access detailed vehicle information.

3. Automated Login and Captcha Solving:

The Vahan Scraping App streamlines the login process by automating the submission of credentials and solving captcha challenges seamlessly. Utilizing sophisticated algorithms and integration with captcha-solving services, the application ensures a smooth and hassle-free login experience for users. By eliminating manual intervention in the authentication process, users can expedite access to the Parivahan website and initiate data retrieval operations without delay.

4. Efficient Data Extraction:

Leveraging advanced web scraping techniques and intelligent data extraction algorithms, the application efficiently retrieves vehicle details from the Parivahan website. Despite the website's limitation of allowing only five vehicle data retrievals per login session, the Vahan Scraping App optimizes resource utilization and maximizes efficiency to ensure seamless data retrieval operations. Through careful management of session persistence and resource allocation, the application circumvents limitations imposed by the Parivahan website, enabling users to access comprehensive vehicle information without constraints.

Key Challenges

The development of the Vahan Scraping App posed several key challenges, among them being the need to design a robust and scalable platform capable of efficiently handling the retrieval of vehicle details from the Parivahan website. The following challenges were identified and addressed during the development process:

1. Automating Captcha Solving:

One of the significant challenges was automating captcha solving, which involved capturing and downloading captcha images from the Parivahan website. We implemented a solution by utilizing a captcha solving API that predicts captcha images, enabling automated captcha resolution and seamless data retrieval from the website.

2. Session Persistence with Stored Cookies:

To maintain seamless user sessions and bypass repeated logins, we implemented a feature to store website login cookies securely in our database upon successful authentication. These cookies were then utilized in subsequent API requests, ensuring continuity of the session until a login was required again. This approach streamlined user interactions and enhanced efficiency by eliminating the need for repetitive authentication steps.

3. Dynamic Browser Instance Management:

To optimize performance and handle multiple requests efficiently, we introduced dynamic browser instance management. This feature dynamically adjusts the number of browser instances opened based on the workload and incoming requests to our API. By dynamically scaling browser instances, we ensured faster processing times and improved responsiveness, enabling the application to handle a higher volume of requests without sacrificing performance.

Applied Technologies

- **Backend:** Flask-Restx
- **Database:** PostgreSQL
- **Load Testing:** Locust
- **Automation & Scraping:** Selenium

Thank You...

INEXTURE

Contact us to get more info

Sales@inexture.com

+91 6353697824

A/B 201-207, Sankalp Iconic Tower,
Opp. Vikram Nagar, Near Iscon Cross Road,
S.G. Highway, Ahmedabad – 380054