Project Overview

The **Email Scraper Web Application** is a Python-based web system developed using the Flask framework. The application automatically extracts publicly available email addresses from websites listed in a CSV file.

It reads URLs from an input CSV file, scans multiple pages (such as Contact and About pages), extracts email addresses using pattern matching and HTML parsing, and stores the results in an output CSV file.

Objective

The objective of this project is to:

- Automate email extraction from multiple websites

- Reduce manual searching effort

- Improve data collection efficiency

- Provide a simple web-based interface for viewing results

Technologies Used

Backend

- **Python**

- **Flask** – Web framework

- **Pandas** – CSV handling & data processing

- **Requests** – HTTP requests

- **BeautifulSoup (bs4)** – HTML parsing

- **Regex (re module)** – Email pattern matching

- **Urllib3** – SSL handling

Frontend

- HTML5

- CSS3

- Jinja2 Template Engine (Flask templating)

Limitations

- Does not scrape JavaScript-rendered dynamic websites

- Only extracts publicly visible emails

- Performance depends on website response time

- No CAPTCHA bypassing

Future Improvements

- Add file upload feature in frontend

- Add download button for output.csv

- Integrate Selenium for dynamic websites

- Add progress bar

- Deploy to cloud (Render / AWS / Heroku)