Puppeteer Scraper API

Problem Statement:

Websites often contain valuable data such as product details, prices, SEO metrics, and performance statistics. Extracting this data manually is time-consuming, error-prone, and inefficient, especially for large-scale data collection or frequent updates. Automating web data extraction enables businesses and developers to make informed decisions, monitor competitors, and optimize digital presence.

Objective:

Develop a Puppeteer-based web scraping API that can programmatically visit websites, extract structured information (e.g., page content, metadata, links, SEO and performance metrics), and return it in a usable format (JSON/CSV). The API should be flexible, scalable, and capable of handling multiple websites while respecting web scraping best practices.

ML/Tech Type:

Web Scraping / Automation / Data Extraction

Input:

Website URL(s), optional selectors or scraping instructions, API key for authentication

Output:

Structured web data in JSON or CSV format, including elements like page title, meta tags, headings, links, images, and performance metrics