Scrape Medicine Details from Pharmaceutical Websites

Domain: Healthcare

Clients: A popular healthcare research firm from Netherlands.

Context: Client wanted to extract medicine details from the catalogs of leading pharmaceutical portals.

Challenge:

Client wanted to extract data for medicines under all the categories available on the pharmaceutical portals at regular intervals to perform analyses over a period of three years. This would result in significant record volume requiring an infrastructure capable of handling big data. The data points that they needed were:

- Disease/Condition Name
- Manufacturer Name
- Salt Name
- Medicine Name
- Price
- Package Size

Solution:

Our team built a custom crawler to regularly extract the required data at scale and deployed it. We also set up monitoring for the target websites in order to promptly update the crawler if the site structure changes. The client was provided with API details and documentation to query and fetch the crawled data.

Benefits to the client:

- The client didn't have to deal with any of the technical aspects in the process
- The setup was completed in just 3 days and the data flow was consistent since then
- We also setup monitoring for the target site to ensure consistent crawling and to avoid data loss
- Our tech stack could efficiently handle the dynamic coding practices used by the target sites
- The client was able start their market research using the delivered data within a short span of time
- The cost of extraction was 78% less than the cost of an in-house crawling setup projected by the client