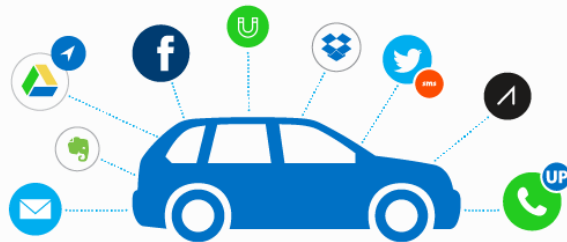# Web Scraping for Cars Data

Domain: Automobile

Solution: Site-specific crawl and extraction

Challenge:

The client wanted cars data to be extracted from automobile websites including classified sites, auto spare parts sites, automobile blogs, forums etc. The required data points were product id, post title, location, seller details, make and model, description and prices.

Solution:

The client provided the list of source websites to be crawled. The data had to be extracted on a daily basis meaning fresh information had to be supplied every day. Our team set up crawlers to fetch the data points from the provided source sites. Since every website in the source list had a different structure and design, site specific crawl and extraction was the offering that suited this case. Once we were done with the initial setup, the crawlers started delivering data. As preferred by the client, the data was to be uploaded to the client's S3 servers in CSV format. Once the setup was complete, our crawlers started delivering about 300 k records on a daily basis.

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
- Our team set up monitoring for the sources to make sure no data was missed
- The complicated cases of dynamic websites were effortlessly handled by our tech stack
- The client was able start their market research using the delivered data within a short span of time
- The cost of data was significantly lower than what the client would incur with an in-house web scraping setup