

# Implementing Web Scraping for Integrating 1.3 Million MRO Products



# Implementing Web Scraping for Integrating 1.3 Million MRO Products

## CLIENT BACKGROUND

- A US Based E-Commerce company wanted to Integrate a comprehensive database of MRO products into their e-commerce platform which includes
- The customer wanted to extract a large dataset of 1.3 million products across 35 categories from a target website.
- Each product category had 400 pages of data & The customer wants to cleanse the data by removing more than 30% duplicate products.
- The target website had over 1000 columns, and the customer wanted images to be downloaded and zipped into a single file, with image links included in the output .csv format.
- And also the customer required a cost-effective and optimized solution for data collection and cleansing.

## Approach To Solution

- The Outsource Bigdata An AIMLEAP Company team developed a custom-programmed web crawler tailored to the client's specific requirements.
- Conducted an in-depth analysis of the target website's structure to identify key data points and relevant pages. Set multiple rules within the crawler to accurately capture product descriptions, specifications, images and all desired data points
- The team automated de-duplication with Python to efficiently remove duplicate products. Developed code to compile and zip images, linking them accurately in the output .csv format.
- The use of custom crawler and automation minimized manual intervention, streamlining data processing for faster, efficient handling and timely delivery of high-quality data.

## Highlight



**75%**  
cost saving



**1.2 million**  
records processed



**99%**  
Above 99% Quality assurance



**Quick turnaround**  
25 business days



## AI & Automation Practice



As part of AIMLEAP Business, AIMLEAP outsource big data practice provides advanced data collection and management expertise, as well as Robotic Process Automation (RPA) capabilities that help clients create highly personalized digital experiences, products and services. Our RPA solutions help customers with insights from data for decision-making, improve operations efficiencies and reduce costs. To learn more, visit us [www.outsourcebigdata.com](http://www.outsourcebigdata.com)

## About AIMLEAP



AIMLEAP - is an ISO 9001:2015 and ISO/IEC 27001:2013 certified global technology consulting and service provider serving small, medium, and enterprise companies helping them to grow with agility to achieve competitive advantage. Our key service offerings include end-to-end IT services, advanced analytics, robotic process automation, web scraping, digital marketing, and BI solutions.

AIMLEAP started its operations in 2012 and successfully delivered digital IT transformation, and data solution projects for more than 550 fast-growing companies in the USA, Europe, New Zealand, Australia, Canada: and more.

- An ISO 9001:2015 and ISO/IEC 27001:2013 certified
- Served 750+ customers
- 12+ Years of industry experience
- 98% Client Retention
- Great Place to Work® Certified
- Global Delivery Centers in the USA, Canada, India & Australia



[outsourcebigdata.com](http://outsourcebigdata.com)



[sales@outsourcebigdata.com](mailto:sales@outsourcebigdata.com)

### USA

80 Broad St, 5th Floor  
PMB#1148, Manhattan, New York  
+1-30235 14656

### CANADA

7030 Woodbine Avenue,  
Markham, L3R 6G2, Ontario, Canada  
+1 437 8370063

### INDIA

Manyata Embassy Tech Park,  
Ground Floor, Beech, E-1,  
Bangalore, Karnataka  
+91-8105 271 615

### AUSTRALIA

21 Hammerwood Avenue  
Derrimut, 3030 Vic, Australia  
+61 402 576 615