**Paper 1**: ESTIMATION OF THE SECOND HAND CAR PRICES FROM DATA EXTRACTED VIA **WEB SCRAPING** TECHNIQUES

**http://jomardpublishing.com/UploadFiles/Files/journals/JTME/V5N2/NasibogluR.pdf**

**Main objectives:**

The main goal of this work is to operate a Web Scraping platform that will collect detailed information related to various vehicle models. from various websites systematically on the internet This document is for the sole purpose of providing quality images of these cars. Along with detailed information such as brand name, specific model, year of manufacture, etc...

**Methodology:**

1. **Data Collection:**

* Car model: Various online channels are used. in car sales Including forums and car marketplaces to retrieve information about different car models.
* Images: High-resolution images of the car model are copied from various angles. to create a detailed image dataset
* Metadata: When it comes to metadata Contains details of each car model as follows:
  + Car brand (e.g. Toyota, Ford)
  + Car model (e.g. Camry, Mustang)
  + Year of construction
  + Engine specifications
  + Color
  + price
  + Additional details such as distance, status, and registration location

1. **Techniques and tools:**

* **Scraper** **Tools**: The programming language used to parse HTML and extract the required data is Python. Some of the libraries include BeautifulSoup and Scrapy, among others, listed below.

**A screen shot of a computer

Description automatically generated**

* **Data management**: To manage content from the web, simply use “requests” to store retrieved data using a structured format, such as post-processing CSV or JSON documents.

1. **Challenges encountered:**

* various sites with a variety of data display formats cause challenges

1. **Solution:**

* Adaptive scraping techniques and data validation methods are used to overcome these challenges and ensure data accuracy.

1. **Data adjustment:**

* Organize collected images and metadata in a central database for efficient access and use to train machine learning models.

**Conclusion:**

The web scraping system effectively gathers a diverse range of car model images and metadata, contributing valuable data for vehicle classification tasks. The methodology demonstrates the ability to handle different website structures and provides a solid foundation for further research and development. Future work will involve expanding data sources and improving scraping techniques to enhance data quality and model performance.

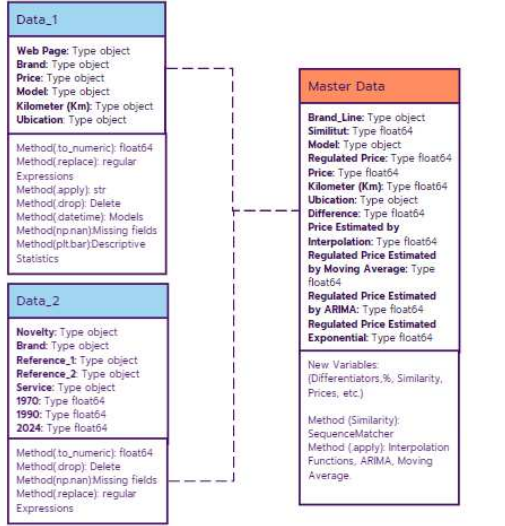
**Paper 2**: Automated Benchmarking for the Purchase of Second-hand Cars and Motorcycles in Colombia through Web Scraping and Prescriptive Analytics

**https://assets-eu.researchsquare.com/files/rs-3693600/v1\_covered\_88b2d2cd-32c3-430a-8325-72053d75c8d9.pdf?c=1702809458**

**The Main Objective**

The main objective of the web scraping phase was to collect detailed and accurate used car information from automotive websites. This includes extracting data points such as:

* Car make
* Model
* Value
* Mileage



The goal was to gather comprehensive data that could be used for subsequent research, including comparing actual prices with regulated prices and providing insights to potential buyers.

**Methodology**

1. **Data Extraction:**

* **Tools used:**
  + **BeautifulSoup**: Used to parse HTML text from web pages and extract content.
  + **Requests**: Used to make HTTP requests to target websites and retrieve HTML data for processing.
* **Approach:**
  + Requests were sent to used car websites.
  + HTML responses were analyzed to extract data on vehicle make, model, price, and mileage.
  + Logic was used to aggregate data from multiple listings across different pages.

1. **Data Cleaning and Pre-processing:**

* **Handling Regular Expressions:**
  + Applied regular expressions to identify and correct irregularities in the data, such as inconsistencies in brand names and formatting issues.
* **Standard setting:**
  + Similar cases were grouped for uniformity and homogeneity in the dataset.
  + Input data were cleaned to remove any discrepancies or inaccuracies.
* **Error correction:**
  + Used predefined rules and patterns to address inaccuracies and standardize data entries.

A diagram of data analysis

Description automatically generated

1. **Integration and Analysis:**

* **Data Modifications:**
  + Data collected from different sources were pooled into a central database for comprehensive analysis.
* **Comparison:**
  + Analyzed the collected data in relation to regulated prices to identify discrepancies and trends.
  + An integrated dataset was used to make comparisons and generate vehicle price insights.

**Conclusion**

* **Effectiveness:** Web scraping efforts yielded robust data for analyzing used car prices. This information is critical for comparing actual prices to regulated prices and providing valuable insights to consumers.
* **Accuracy:** The use of data cleaning techniques and tools such as BeautifulSoup and Requests made the extracted information more accurate and reliable.
* **Challenges:**
  + Addressed issues related to inconsistent data entries and formatting errors through effective cleaning and preprocessing methods.