

# Optimal Product Price Analyzing System

Akshat Goel  
SRM Institute of Science &  
Technology, Tamil Nadu

Sachin Verma  
SRM Institute of Science &  
Technology, Tamil Nadu

Dr. Prabakeran S.  
Department of Networking and  
Communications  
SRM Institute of Science &  
Technology, Tamil Nadu

**ABSTRACT:** Online Shopping has become the standard, due to the rapid increase in the number of e-commerce websites. While internet buying is convenient, determining which website has the best pricing and offers is a time-consuming and laborious effort. For a shopper, comparing products and screening them from each internet site takes a long time. This paper uses Web extraction techniques using Python modules/libraries like Beautifull Soup, requests, and matplotlib library to get the best review rates and the best offer on product for the end-user from a variety of marketplace or e-commerce websites. Web extraction is a way of obtaining raw/unstructured data from web pages with the help of a computer. In this paper, raw/ unstructured data is taken from a variety of marketplace websites and compared automatically. Finally, based on the customer's appropriate selection, the findings are visually shown.

## I. INTRODUCTION

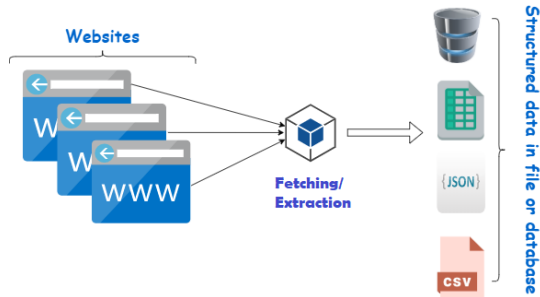
The optimal lower price analysing System is used to locate the lower price available of a product from Ecommerce websites such as amazon, Flipkart, Snapdeal, paaytm, CRroma, and others. It entails the use of cutting-edge technologies like online scripting, data science, and data analytics. We will be able to not only identify the lowest price of a product, but also analyse the product data available on the internet, by using the ptimal ost roduct rice nalyzing System. It includes a variety of critical procedures such as analysing, modelling, visualising,

and evaluating. It has a wide range of application possibilities, including price monitoring, news monitoring, trend recognition, and lead generation. When looking for information on the internet, most people utilise a web browser.

Browsers make viewing and accessing various websites simple and straightforward. Large volumes of data in an unstructured format are found on websites. There is a lot of garbage/ junk mixed in with raw material on a web page. Relevant data extraction must be completed in order to extract useful & meaningful information from a webpage. This can be performed by employing webscanning tools. "Web scrapping" is a technique for obtaining data from websites. Web screping is also called as "web harvesting," "web data extraction," and "web data mining." It's a scheduled & automated method for swiftly and easily retrieving big amounts of data.

Massive Data/ information is extracted, filtered and stored in a structured manner (such as .XLS or .CSV files or databases). The legality of using online scripting is entirely up to the website administrators. People will be able to access the data on a specific website if they consent. Data is transformed from an unstructured to a structured format via Web extraction. Online businesses are now easy to launch, and have inexpensive initial expenses all around the world. People prefer to create an online store for a variety of reasons, including reduced taxes, fewer crowds, greater selection, and quicker updates. Customers are more inclined to use e-commerce services as the number of them expands to spent a significant amount of money, time on price, review rating, product characteristics, and delivery time.

This article explains one way for pulling data from e-commerce websites and showing it on a customer's screen, allowing them to sort through a big number of unnecessary information. Python, Node.js, H, Ruby, C++, and other programming languages can be used to implement web cleaning.



## II. LITERATURE SURVEY

Renita rystal ereira provided web scrapping summary and techniques and tools that face several complexities as data extraction isn't that simple. These tactics ensure that the data collected is correct, consistent, and has higher integrity because there is a large amount of data that is difficult to handle and keep. Though there are certain issues with functional techniques, such as the increased amount of web scrapping being able to cause severe harm to websites. When it comes to stealing information, this also provides hackers an advantage. In terms of business, social networking is crucial because it is here that the concept of rising income originates. It will also help consumers save time and money by allowing them to shop quickly, similar to online shopping. Supporting and earning from a company, on the other hand, has its advantages. Kaushal Arikh et al. [2] advocated utilising machine learning to detect web scrapping. It's useful for businesses that rely on research. For a long time, web scrapping has been a challenging preventive attack.

When a business publishes information on the internet, it is possible that it will be copied, pasted, and then utilised in another context without the company's knowledge. As a result, web mapping

solutions have a big influence on the conclusion of the cause. Sameer Adghan et al., [3] projected a approach for easily extracting data from internet pages in conjunction with web scrapping.

This technology would allow data to be scraped from a wide range of websites, reducing human interaction, saving time, and boosting data relevancy quality. It will also help the user collect data from the site, save it according to their needs, and use it as they see fit. The scribbled information can be used for database construction, research, and other related tasks. This strategy should be regarded as a gift that must be managed with care in order for human races to progress. The most recent technique, known as Web Scrapping, was found by Nand Saurkar et al., [4].

Web scrubbing is a critical approach for converting unstructured data on the internet into structured data. Scrapping formed structural data, which was then gathered and evaluated in spreadsheets in the central database. The writers discussed the essentials of web processing in this essay. They concentrated on web scrapping techniques. The study concludes with a survey of the different technological resources for effective online scrapping that are currently available on the market. [5] Federico Olidoro et al. focused on the outcomes of web scrapping evaluation strategies with particular orientatiton to user electronics services and goods throughout the sector.

Although the research done has so far been performed in a small amount of time, you can see in whatever followed, it has enabled to attain important, but not conclusive. This topic has been briefly addressed in the points assigned to both of the examined items, but in reality, interacting with this viewpoint necessitates a consideration of the current survey architecture, which does not require or only selectively permits the use of large data approaches within the existing sample frameworkK. Jan Kinne et al. proposed a web-based platform for the accurate and quantifiable mining of ecosystems for development.

The main purpose of the suggested method is to find and extract specific bits of data from the site's

unstructured content that reveals information about organisations' present development practises. A technology like online scrapping can emerge, according to Ingolf Boettcher [7].

Web scrapping innovation offers a variety of options and can be used for a variety of purposes: A web crawler's primary purpose is to automate the previously physical task of obtaining pricing estimations and online article details. Any use of the approach demands rigorous planning in several areas. First and foremost, ethical and data security concerns must be addressed. Within the context of a research project, essential IT services and IT practise necessary to administer the automated information gathering system must be calculated and not disregarded.

[8] The measurement level of the online scrapper will differ from the measurement units of the original source file, making it impossible to comprehend the data. The number of people utilising social networking sites on the internet, such as Facebook, Twitter, LinkedIn, and others, is increasing every day. The internet, which is available to everyone, amplifies knowledge. [9] Many protection systems have already been put in place, yet some are still being ignored. The importance of machine learning increases as a result.

Machine learning is quite good at detecting patterns. As a result, if we can successfully teach the system to recognise an intruder's cadences, we will be able to avoid such threats. Scrapping solutions translate complex data obtained through networks into structured data that can be stored and examined in a central dat [10]. The amount of scrapping used would skyrocket, and it would frequently encroach on the framework to obtain the information. Scrapping, on the other hand, can be prevented by using effective and secure online scrapping techniques.

This study looks at a summary of the web scraping data extraction process, a few web scraping strategies, and the majority of the most recent web scraping tools. This methodology's main purpose was to collect web-based data and incorporate it into a specified repository. [11] Web scrapping tactics

used in the growth analysis will expose the researcher to a larger volume of data than is now available in the data set, perhaps increasing the growth estimate. A web crawler's ultimate purpose is to uncover previously inaccessible pricing data outlets and construct a list of all pricing data that is available online. The actions to build web scrapping for price analytics include sizable analytics and administrative consequences.

[12] While examining technology structures across corporate websites to see if all of these sorts of companies could be measured using the suggested method, researchers focused a special emphasis on uncovering a probable bias. As a research tool, web extraction still needs to deal with enormously huge and unconnected external websites, and restricted broadband access continues to prevent enterprises from managing their internal websites and therefore avoiding web mining research. The suggested research method allows for the integrated, costeffective modelling of complete business communities, which can be done more efficiently and in substantially less time than traditional methods. To simulate information communities, this strategy can be readily extended by visiting the web pages of research institutes.

### III. METHODS

- **Python** is an Object Oriented Language, its also a Functional & Procedural language. It considers to be the best language as it comes with libraries & modules. It's a multipurpose language with a fantastic interactive environment.
- **Requests** module in Python is used to create HTTP requests for retrieving web elements. We can obtain a webpage's raw HTML element, which we will use to extract information from.
- **Beautiful Soup** is a high-level web scraping package that is used to download in Python interpreter by typing "pip install beautifulsoup4" at the console. It uses webpage & provides an interface to navigate the elements. It is also used in fetching data out of the XML & HTML files.
- **Matplotlib** is a Python visualisation package for 2D plots of data arrays that is free, open-source,

and friendly. It's a NumPy array-based multi-platform data visualisation framework. One of the most important advantages of visualisation is that it sanctions us to see immensely colossal magnitudes of information in very easy & understandable format. It also includes a number of other tools that extend Matplots' capabilities.

## IV. WORK FLOW

### A. Obtaining information from web elements

Collect the URL/link of all the products from different marketplace for extraction of data. The link of the products needs to be checked whether they are live or not. And, if the link is live we can make request to the webpage and can extract the content from the webpages.

### B. Extraction of Information

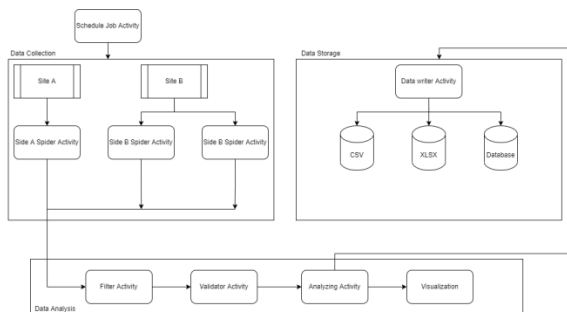
The downloaded resources are in HTML format and are primarily unstructured. As a result, the BeautifulSoup module is used to fetch and extract structured data from the downloaded resources.

### C. Managing the Information

The raw data which is extracted from the webpage is needed to be stored in database for further analyses & visualization. We can store data in the form of CSV, XLS or JSON.

### D. Data Analyzing & Visualization

The extracted data that is stored will be analysed by the web scraper. By developing code using various Python libraries, the stored data can be changed.



## V. IMPLEMENTATION

1. For our research, we use two marketplace/ e-commerce websites for extracting product prices such as mobile phones, electronic gadgets such as camera, printer, charger etc. We need to find the Xpath of the data which we need to extract. We can do so by inspecting the web page by right clicking on the page and go to inspect element and the then copy Xpath.
2. We need to look into the web page to identify which element contains the information we want to extract. We use Python to scrape data from a marketplace. We need to import some modules such as Requests and BeautifulSoup, as well as additional data analytic libraries such as numpy, Pandas & matplotlib.
3. The Requests Library/ module then generate HTTP requests to fetch the webpage link for the extraction of raw & unstructured data from the web elements. The HTML data is then parsed using the BeautifulSoup package. Now, the structured data can be analyzed and visualised with the help of matplotlib module and them it can be stored in database or in CSV or XLS format.
4. Because the unstructured data is formatted and stored in a local data base or CSV file, it may be accessed easily at any time. Any e-commerce site's data is constantly changing, which means that product pricing can change at any time, and certain items could be unavailable. This periodic change in data should be factored accordingly in the code.
5. Big chunks of data are dynamically, fetched, extracted and saved in CSV format. For visualisation of the retrieved data, the information can be stored in CSV or XLS format and plotted using the matplotlib tool in Python.

## VI. RESULTS

We have extracted the price, rating and rating count of the product from marketplaces like Amazon and Flipkart. Below table shows the extracted and filtered data from product WebPages.

Product	Amazon_Price	Amazon_Rating	Amazon_Rating_count	Flipkart_price	Flipkart_Rating	Flipkart_Rating_count
Redmi 9 Activ (Coral Green, 4GB RAM, 64GB Storage)	9999	4.2	148244	10295	4.2	1728
Redmi Note 10 Pro (Glacial Blue, 6GB RAM, 128GB Storage)	17499	4.2	54790	17824	4.2	4332
Redmi Note 11 (Horizon Blue, 4GB RAM, 64GB Storage)	13499	4	2072	14170	4.3	38
Redmi 8A Dual (Sky White, 2GB RAM, 32GB Storage)	8199	4.1	81166	8561	4.2	17596
Redmi Note 10T 5G (Graphite Black, 6GB RAM, 128GB Storage)	15999	4.1	8677	16558	4.2	827
Samsung Galaxy M12 (Blue, 4GB RAM, 64GB Storage)	9999	4.1	124825	10510	4.2	10127
Samsung Galaxy M32 (Black, 4GB RAM, 64GB Storage)	14999	4.2	32665	14441	4.2	1378
Samsung Galaxy A12 (Black, 4GB RAM, 64GB Storage)	12999	4	1727	12999	4.2	3850
Samsung Galaxy F12 (Celestial Black, 4GB RAM, 64GB Storage)	11999	4.1	30	9799	4.3	232565
Samsung Galaxy M11 (Black, 4GB RAM, 64GB Storage)	12498	4.1	23760	11998	4.2	9611

After extracting the data we filtered it even more based on price and rating. For rating count, we cannot directly filter based on rating or rating count. So, instead we use multiplication of rating and rating count of respective marketplace. So, below tables show the filtered data based on price and rating.

### FILTERED DATA BASED ON PRICE

Product	Best_Price	Marketplace
Redmi 9 Activ (Coral Green, 4GB RAM, 64GB Storage)	9999	Amazon
Redmi Note 10 Pro (Glacial Blue, 6GB RAM, 128GB Storage)	17499	Amazon
Redmi Note 11 (Horizon Blue, 4GB RAM, 64GB Storage)	13499	Amazon
Redmi 8A Dual (Sky White, 2GB RAM, 32GB Storage)	8199	Amazon
Redmi Note 10T 5G (Graphite Black, 6GB RAM, 128GB Storage)	15999	Amazon
Samsung Galaxy M12 (Blue, 4GB RAM, 64GB Storage)	9999	Amazon
Samsung Galaxy M32 (Black, 4GB RAM, 64GB Storage)	14441	Flipkart
Samsung Galaxy A12 (Black, 4GB RAM, 64GB Storage)	12999	Flipkart
Samsung Galaxy F12 (Celestial Black, 4GB RAM, 64GB Storage)	9799	Flipkart
Samsung Galaxy M11 (Black, 4GB RAM, 64GB Storage)	11998	Flipkart

### FILTERED DATA BASED ON RATING

Product	Best_Rating_Marketplace
Redmi 9 Activ (Coral Green, 4GB RAM, 64GB Storage)	Amazon
Redmi Note 10 Pro (Glacial Blue, 6GB RAM, 128GB Storage)	Amazon
Redmi Note 11 (Horizon Blue, 4GB RAM, 64GB Storage)	Amazon
Redmi 8A Dual (Sky White, 2GB RAM, 32GB Storage)	Amazon
Redmi Note 10T 5G (Graphite Black, 6GB RAM, 128GB Storage)	Amazon
Samsung Galaxy M12 (Blue, 4GB RAM, 64GB Storage)	Amazon
Samsung Galaxy M32 (Black, 4GB RAM, 64GB Storage)	Amazon
Samsung Galaxy A12 (Black, 4GB RAM, 64GB Storage)	Flipkart
Samsung Galaxy F12 (Celestial Black, 4GB RAM, 64GB Storage)	Flipkart
Samsung Galaxy M11 (Black, 4GB RAM, 64GB Storage)	Flipkart

## VII. CONCLUSION

This paper provides a revolutionary collaborative Price filtering method for shopping sites based on Web scraping techniques. A relevant case study of analysis of price & comparison is done from marketplace or e-commerce websites, is given to show how the system works in practice. Compare prices on many web sources, if possible. Data can be scraped from a variety of sources, including social media, employment boards, and travel websites. The amount of data available on the internet is steadily expanding. Now this data can be stored in local data servers or computer for ease of retrieval. Web extraction and analysis of data will play a critical part in data extraction in the future due to the vast volumes of data available on the internet. Scrappers, on the other hand, should keep in mind that they are not infringing any laws that could put them in legal trouble.

Future enhancement that can be made:-

- By using multiple proxies we can distribute the requests made from user-end.
- By using a scheduler we can schedule our requests at particular time & interval, this also prevent our requests from getting blocked.
- By using cloud servers we can automate our bot to work at any particular time & interval.

## VIII. REFERENCES

- [1] Renita Crestal Pereira and T.Vanita, "Online Extraction of Media Websites," International Innovative Research Journal in Com. and Comm. Engineering, pp. 240-276, Vol. 3, 2015.
- [2] Fedireco Polodiro, Ricardo Gianini, Rosana Lonte, Stafeno Monsca & Francesca Rosetti, "Online Extraction ways to fetch info. on Commercial hardware and airlines for American

HCIP,” Statisstical Journal of the IOAS, pp. 156-167, 2015.

- [3] Kaushal Parikh, Dilip Singh, Dinesh Kumar and Mansingh Rathode, “Machine Intelligence in Online Extraction,” KGMP international journal of Science and Information Technology, pp.141-150, Vol. 3, 2018.
- [4] Ananda V.Saurkar, G.Kedar & Sheweta Gowde, “Detailed Study on Web filtration & Storage Access,” Open Access National Index of Science & Information Engineering, pp. 336-341, Vol. 5, 2018.
- [5] Sameer Padghan, Satish Chigle and Rahul Handoo, “Online Extraction of Media Websites,” International Innovative Research Journal in Com. and Comm. Engineering, pp. 692-696, Vol.16, 2018.
- [6] Jane Kine and Janna Axanbeck, “Online Minning of Corporate Web Pages”, International Innovative Research Journal in Com. and Comm. Engineering, 2019.
- [7] Eric J.Farley and Lissa Pierote, “An evolving information analyzing Method for Criminal & Law Justice Researchers,” Justice Analytics Research & statistical association, pp. 2-10, 2020.
- [8] Mitchal, Rayn. “Online Extraction with .py: Collecting data available from the modern web pages. "O'Reilaly Plat., Inc., 2018.
- [9] F.Zhao, J.Zhou, C.Nie, H.Huang, and A.Jin, (2016) IndexCrawler: A Three-stage spider & crawler for valuable Harvesting Deep online Interfaces, IEEE transections on computing services.
- [10] Chaulagain, Ram Sharan, Subrna Shakrya. "Cloud Computing based online scraping for big-data applications." In 2017 IEEE International Conference on SmartCloud (SmartCloud), pp. 139-144. IEEE, (2017).
- [11] Madhusudan, Lumbhate Ponam, D. (2018) Deep-Web Crawling Efficiently using static based python bot Crawlers, International Journal of Computer Science & Engineering (IRJET), 04(07), pp. 3304.
- [12] Mitchal, Rayn. “Web Extraction & Valuation using Python: Evaluating data

available from the marketplace websites.  
"O'Reilaly Plat., Inc., 2019.