Automated Web Scraping and Data Visualization from Multiple ECommerce Websites Using UiPath

[1]Nisha A, [2]Mandalam chitralekha

[1]M.E., CSE PG Student, Department of Computer Science and Engineering, SRM Valliammai Engineering College.

[2]Assistant Professor, Department of Computer Science and Engineering, SRM Valliammai Engineering College.

[1]nishaanandan123@gmail.com, [2]chitralekha230401@gmail.com

Abstract-**In the digital era, online shopping platforms like Amazon, Flipkart, and Croma offer vast product catalogs with dynamic pricing and customer ratings. Extracting and analyzing this data manually is inefficient and time-consuming. This project focuses on automating web scraping using UiPath to extract essential product details such as name, price, and ratings from multiple e-commerce websites. The scraped data is then exported into Excel or CSV format for further analysis. Using visualization tools, key insights are derived from the collected data through graphs and charts, enabling better decisionmaking for market analysis, price comparison, and consumer trends. This study demonstrates the power of automation and data analytics in e-commerce research and competitive analysis.**

Keywords— Web Scraping, Ui-Path, Automation.

I. INTRODUCTION

RPA(Robotic Process Automation) :

Robotic Process Automation (RPA) is a technology that automates repetitive, rule-based business tasks using software robots, or bots procedures. By resembling human behaviors like clicking, typing, and interacting with apps, these bots can increase efficiency and decrease mistakes.

Ui-Path :

UiPath is one of the leading RPA platforms that provides a user-friendly interface for automating tasks. It offers:

- **UiPath Studio** (for designing workflows)
- **UiPath Orchestrator** (for managing bots)
- **UiPath Robot** (for executing automation tasks)

The project involves a systematic approach to extracting and analyzing e-commerce data. The key steps include website identification, web scraping using UiPath, data storage and cleaning, data preprocessing, data visualization, and decisionmaking insights. The first step is selecting target websites such as Amazon, Flipkart, and Croma.UiPath is then used to automate data extraction for product details like name, price, and ratings. The extracted data is stored in Excel/CSV files and cleaned for better analysis. Preprocessing includes handling missing values, filtering irrelevant data, and structuring the dataset. Data visualization follows, where meaningful insights are generated using graphs and charts to identify pricing trends and market patterns. Finally, the interpreted data helps businesses make strategic decisions related to competitive intelligence and consumer behavior analysis.

II. LITERATURE REVIEW

A literature review is a comprehensive analysis of existing research on a specific topic, identifying key themes, gaps, and methodologies. This study examines pricing strategies like costplus, value-based, penetration, and premium pricing, highlighting their impact on businesses and consumers. It explores market influences, psychological effects, and AI-driven dynamic pricing while addressing challenges in balancing profitability and competition.

This paper explores web scraping techniques for extracting and analyzing online data, emphasizing applications in market research, price tracking, and sentiment analysis. It discusses Python libraries like BeautifulSoup, Scrapy, and Selenium, along with data cleaning and visualization methods. Ethical considerations and real-world uses, such as dynamic pricing and consumer trend analysis, are also examined.

This paper examines how web scraping aids in real-time price comparison across e-commerce platforms, helping consumers and businesses optimize decisions. It discusses Python tools like BeautifulSoup, Scrapy, and Selenium for extracting and analyzing pricing data while addressing challenges like website restrictions and data.

This study proposes the development and implementation of an automated web scraping solution using **Robotic Process Automation (RPA)** with the **UiPath** platform. The primary objective is to automate the extraction of structured and semistructured data from web sources, reducing manual effort and enhancing the efficiency, accuracy, and scalability of data collection processes.

III. SYSTEM DESIGN

The suggested system will make use of UiPath's built-in features, including Custom Selectors, UI Automation, and the Data Scraping Wizard, to browse web pages and extract pertinent data. AJAX-based loading mechanisms, dynamic content, and multi-page navigation are just a few of the website structures that the automation workflow will be able to manage.

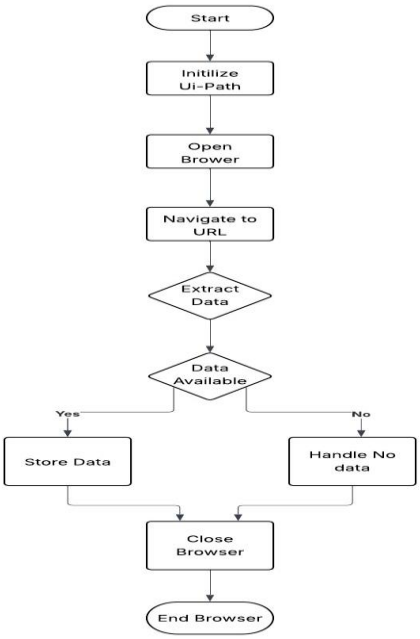

Fig. Flow Diagram for Ui Path (RPA) work flow for proposed work

IV.SYSTEM IMPLEMENTATION AND RESULT

A General UiPath workflow for extracting product name, price, rating, and URL:

1. **Open Browser/Application:** The workflow typically starts by launching the web browser or desktop application containing the product listings using activities like "Open Browser" or "Open Application."
2. **Navigate to Target Page:** Activities like "Go To URL" or UI interaction activities (e.g., "Click") are used to navigate to the specific page(s) with the product details.
3. **Identify Product Containers:** UiPath uses UI selectors to identify the repeating elements that contain individual product information (e.g., a div or list item for each product).
4. **Data Scraping/Extraction:** The "Extract Data From Table" or "Data Scraping" wizard is often employed to define patterns for extracting the product name, price, and rating from within each identified product container. For the URL, you might need to extract an attribute (like "href") from an anchor tag associated with the product.
5. **Iterate and Extract (if needed):** If product details are on multiple pages, a "While" loop or similar iteration logic along with "Click" activities for "Next" buttons would be used to process all pages.
6. **Store Extracted Data:** Typically, UiPath stores the extracted data in a DataTable variable with the columns "Product Name," "Price," "Rating," and "URL."
7. **Write to Output:** Finally, activities like "Write Range" (for Excel) or "Write CSV File" are used to output the collected data into the desired format.

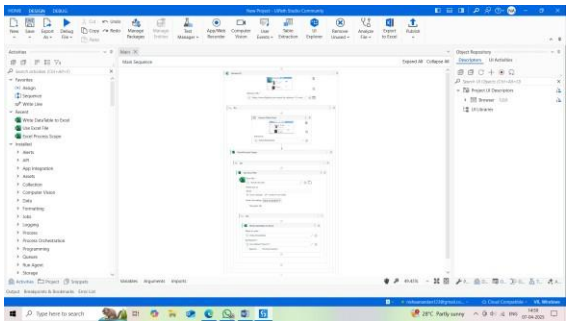

FIG.2 Ui-Path work flow

The above figure denotes that UiPath workflows are designed to automate repetitive tasks and streamline business processes. They use a combination of sequences, flowcharts, and state machines to structure automation efficiently. UiPath Studio offers a number of activities for handling exceptions, interacting with applications, and manipulating data. Web scraping and data analysis can be accomplished by integrating Python libraries such as BeautifulSoup, Scrapy, and Selenium.

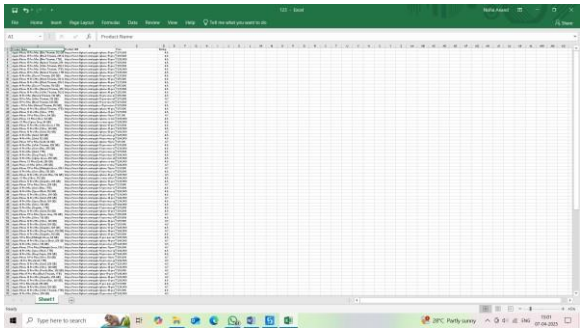

Fig.3 Extracted data in Excel

This content, when used within a UiPath workflow, would typically be the *output* of a data extraction process. Here's how UiPath would generally handle this:

1.Data Extraction Activities: UiPath offers various activities to extract data from different sources:

- **Read Range (Excel):** If the data already exists in an Excel file, this activity reads it into a DataTable.
- **Extract Data From Table (Web/Desktop):** If the data is on a website or in a desktop application in a tabular format, this activity can scrape it.
- **Get Text/Get Attribute:** For more unstructured data, these activities can extract specific pieces of information based on selectors.
- **Computer Vision Activities:** For scenarios where traditional selectors are unreliable, computer vision can be used to identify and extract data based on visual elements.

2. Data Storage (DataTable): Once the data is extracted, UiPath usually stores it in a data structure called a DataTable. This is a temporary, in-memory representation of tabular data.

3.Writing to Excel: To create the output Excel file, UiPath uses activities like:

- **Write Range (Excel):** This activity takes a DataTable as input and writes its contents to a specified Excel file and sheet. If the file doesn't exist, it will be created.

V. CONCLUSION AND FUTURE WORK

In conclusion, the UiPath RPA workflow successfully automates product data extraction (name, price, rating, URL) and saves it to Excel, eliminating manual effort and providing a valuable dataset. Future work will focus on enhancing robustness against dynamic websites, automating navigation across multiple pages, implementing data cleaning and transformation, expanding the scope of extracted details, enabling scheduled runs and centralized management, improving error handling with notifications, integrating with other business systems, leveraging AI for complex data extraction, optimizing performance, and ensuring ethical web scraping practices.

REFERENCES

[1] Vayadande, K., Bodhe, Y., Bhosle, A., Desai, P., Dere, G., & Elkiwar, S, "An Automated Approach to Finding the Best Prices on GeM and Private E-Commerce Sites", Nanotechnology Perceptions, 20(S6), 829-844, 2024.

[2] Bi Bi Hajira Khanum & Dr. Raghavendra S, "Price Comparison using Web-scrapping and Data Analysis", 2024, https: www.doi.org/10.37314/JJEM.SP0256.

[3] V. Srividhya, P.Megala "Scraping and Visualization of Product Data from E-commerce Websites", 2024 https://doi.org/10.26438/ijcse/v7i5.14031407.

[4] Mehra, A., Sharma, S., & Kaoushal, V. "Implementation of government e-marketplace in tertiary care hospitals of North India: Advantages and challenges", International Journal of Research in Medical Sciences, 2020.

[5] Abodayesh, A., Hajazi, R., Najjar, W., Shihadeh, L., & Latif, R, "Web scraping for data analytics: A Beautiful Soup implementation" Sixth International Conference of Women in Data Science at Prince Sultan University (WIDS PSU), 2023.

[6] Khder, M. A., "Web scraping or web crawling: State of art, techniques, approaches and applications". International Journal of SoftComputer Application, 2020.

[7] Singrodia, V., Mitra, A., & Paul, S, "A review on web scraping and its applications", International Conference on Computer Communication and Informatics, 2021.

[8] Elhussein, B., & Karrar, A. E, "Review the Amazon products using data science tools to enhance online feedback and services", Second International Conference on Smart Technologies Smart Nation (SmartTechCon), 2023.

[9] Eloisa Vargiu, Mirko Urru, ―Exploiting web scraping in a collaborative filteringbased approach to web advertising‖, Artificial Intelligence Research, Vol. 2, Issue 1, pp. 44-54, 2013.

[10] Vasani Krunal A., ―Content Evocation Using Web Scraping and Semantic Illustration‖, IOSR Journal of Computer Engineering
(IOSR JCE) Vol. 16, Issue 3, pp. 54-60, 2014

[11] Jose Ignacio Fern ´ andez-Villamor, Jacobo Blasco-Garc ´ ´ıa, Carlos A. Iglesias, Mercedes Garijo, ―A Semantic Scraping Model for Web Resources-Applying Linked Data to Web Page Screen" in the Proceedings of the 3rd International Conference on Agents and Artificial Intelligence, Volume 2 - Agents, Rome, Italy, 2011.

[12] Emilio Ferraraa, Pasquale De Meob, Giacomo Fiumarac, Robert Baumgartner, ―Web Data Extraction, Applications and Techniques: A Survey‖, Elsevier, knowledge-Based Systems, pp. 301-323, 2014.

[13] Faustina Johnson and Santosh Kumar Gupta. ―Web Content Mining Techniques: A Survey‖, International Journal of Computer Applications Vol. 47, Issue 11, pp. 44-50, 2012.

[14] Rahul Dhawani, Mrudav Shukla, Priyanka Puvar, Bhagirath Prajapati, "A Novel Approach to Web Scraping Technology‖, International Journal of Advanced Research in Computer Science
and Software Engineering, Vol. 5, Issue 5, pp.747-750, 2015

[15] Anand V.Saukar, Kedar G.Pathare, Shweta A. Gode, ―An Overview on Web Scrapping Techniques and Tools", International Journal on Future Revolution in Compuer Science & Communication Engineering, Vol. 4, Issue 4, 2018.

[16] Riya Shah, Karishma Pathan, Anand Masurkar, Shweta Rewatkar, Prof. (Ms.) P.N.Vengurlekar, ―Comparison of E-commerce Products using web mining‖, International Journal of Scientific and Research Publications, Vol. 6, Issue 5,pp.640-644, 2016.

[17] Nakash, J., Anas, S., Ahmad, S. M., Azam, A. M., Khan, T. ―Real Time Product Analysis using Data Mining" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol. 4, Issue 3, pp. 815–820, 2015.

[18] Rehman S.U., "Smart agent for automated E-commerce" 2011 World Congress on Sustainable Technologies (WCST), UK IEEE pp.124-128, 2011.

[19] Shikha Mahajan, Nikhit Kumar , ―A Web Scraping Approach in Node.js‖, International Journal of Science, Engineering and Technology Research (IJSETR) Vol. 4, Issue 4, pp. 909-912, 2015.

[20] Sneh Nain, Bhumika Lall, ―Web Data Scraper Tools: Survey‖, International Journal of Computer Science and Engineering, Vol. 2, Issue 5, pp. 39-44, 2014.

A.Nisha was born in Guduvancheri, Tamil Nadu, India, on March 23, 2003. She received the B.E. degree in Cyber Security from SRM Valliammai Engineering
College, Chennai, India, in 2024. She is currently pursuing the M.E. degree in Computer Science and Engineering at SRM Valliammai Engineering College, Chennai, India. Her major field of study is Software Engineering and Cloud Computing, Data science, Machine Learning. She has completed internships in Penetration Testing, Junior Security Analyst, Basics Of Cyber Security. During her undergraduate studies, She worked on a project titled Developing a automation malware detection tool using machine learning algorithm. Her current areas of interest include machine learning, and data science. A.Nisha has participated in academic and project-based activities throughout his academic career. He actively contributes to research discussions and technical forums within his institution.

Ms. Mandalam Chitralekha is a Research Scholar in the field of Information and Communication Engineering. She
Completed her M.Tech in Data Science from Anna University, Rajalakshmi Engineering college, Chennai (2024), and B.E. in Electronics and Communication Engineering from Anna University, GRT Institute of Engineering and Technology, Tiruttani (2022). Area of interest includes the development of new technologies using Data Science Algorithms, Big Data Analytics, Genomics and Robotic Process Automation (RPA).