

Automated Web Scraping System for E-Commerce Data Extraction Using Python

Taniya Pawan Pal

Department of Science and Technology, G.H.Raisoni Skill Tech University, Nagpur, India

ABSTRACT

We live in a world where we buy things online. There are many websites that sell things and they have a lot of details about the products they sell. These details include the price of the products what other people think of them and what the products are. It takes a time to look at all these details by hand. That is why we use something called web scraping to collect details from websites. This project is about a system that uses Python to collect details from Flipkart. The system uses tools like Requests and BeautifulSoup and Selenium to get details about the products on Flipkart. We want to know the names of the products how much they cost what people think of them and what they are. We take the details we collect. Make them clean and easy to use. We store the Flipkart details in a format like a CSV file or a SQL database. We can use the Flipkart details we collect to see what is happening in the market. This system shows how we can use automation and collecting Flipkart details to help us make decisions and to do research on Flipkart products. The system is designed to extract product details from Flipkart. It uses Python to do this. We use the collected Flipkart data to do things like market analysis, price comparison and consumer behavior analysis. Consumer behavior analysis is the study of what people buy from Flipkart and why they buy these products from Flipkart. This project is about using automation to collect details from websites like Flipkart. It shows how this can be useful, for people who want to do research or make business decisions using Flipkart details. We can learn a lot from the details we collect from Flipkart. We can use Flipkart details to make decisions.

Keywords: Web Scraping, Python, Flipkart, Data Extraction, BeautifulSoup, E-commerce Analysis, Data Mining.



This is an open-access article under the [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) license

I. Introduction

Shopping is popular nowadays. We find a lot of details about products, prices and ratings online. We see what people think about these products. Websites like Flipkart have lots of product information.

- ✓ We can use this info to learn about market trends
- ✓ We can make product decisions using Flipkart.

Collecting info from websites ourselves takes time. It's a lot of work because it takes time. So we use web scraping to get info from Flipkart. Web scraping helps get info from websites like Flipkart. We use it to get product info from Flipkart. We can use a program like Python to get product info from Flipkart. This way we store product info from Flipkart in a to-understand format. This project is

about creating a system that automatically gets product info from Flipkart. The system stores product info from Flipkart in a to-read format. We want to get product info from Flipkart. Then we use this info to make charts and pictures. These charts and pictures help us understand the product info from Flipkart. We will use Flipkart to get product info. This info helps us make product decisions. We get product info from Flipkart. We use it to learn more about products, on Flipkart.

Proposed AI-Based Solution

The system they are talking about uses Python to get product information from Flipkart. This system sends requests to the Flipkart website. Looks at how the website is set up. Then it pulls out the details about each product.

The system gets details like the product name. It gets the price of the product. It gets what people think of the product. It gets what people say about the product. It gets what the product is about.

- ✓ The product name
- ✓ The price of the product
- ✓ What people think of the product
- ✓ What people say about the product
- ✓ What the product is about

Then the system makes sure all the information about the product is correct. It puts the information about the product in order. The system saves the information about the product in a way that's easy to use. For example it saves the information about the product in a CSV file or a SQL database. Using Python to get product information from Flipkart saves a lot of time and work. The system can collect a lot of information about products quickly. This is really helpful when you need to look at a lot of products at the time. You can get information about products from Flipkart quickly. This is really helpful when you need information about products, from Flipkart.

Motivation

We are working on this project because people need information from stores like Amazon or eBay. They want to know what's available on these websites. For example people want to know what's popular on Amazon or eBay. They also want to know how often people buy these products from Amazon or eBay.

- The problem is that many websites do not make it easy to get this information from Amazon or eBay.

This information is helpful for people who study how people shop on the internet. They want to understand how people behave when they shop online on Amazon or eBay. Our project helps with that. It is also helpful for understanding what products people like to buy from Amazon or eBay. We use Amazon or eBay and web scraping to get product information. This helps with studying how people shop on stores like Amazon or eBay. We are using web scraping, on Amazon or eBay to help people. It helps businesses and researchers understand Amazon or eBay. They study people who shop on Amazon or eBay.

II. Related Work

People have looked into using web scraping to get data from the internet. Analyze it. For example Mitchell in 2018 said that you can use Python libraries like BeautifulSoup and Scrapy to get data from websites.

- **Lawson back in 2015 talked about how web scraping can be used for real business and research.**

- **Marrese-Taylor and others in 2014 did a study on web scraping and they found out that it is really important for getting structured data from web pages.**

All these studies show that web scraping is really important for finding data looking at markets and doing research. Web scraping plays a role, in these things and it is very useful for web scraping to get the data we need from websites.

III. Research Methodology

This project uses a method to do its research and it has several steps.

- ✓ The research methodology used in this project includes things like
- ✓ Looking at the Flipkart website to see how it is set up.
- ✓ Finding the product data fields that're important.
- ✓ Making a tool with Python to get information from the website.
- ✓ Getting product data from pages, on the Flipkart website.
- ✓ Cleaning up the data that we got and making it nice and tidy.
- ✓ Putting the data in databases that make sense.
- ✓ Doing some work to understand the data that we got from Flipkart.

Phases Phases of Web Scraping System

1. Requirement Analysis Phase

The Web Scraping System should find product details on websites. It should find things like the product name and price. It should also find reviews and comments about the product. We need the Web Scraping System to get all this information about a product. The Web Scraping System has to find out what people think about a product.

2. System Design Phase

Next we plan how the Web Scraping System will work. The Web Scraping System has parts that work together. These parts include getting data from websites making sense of the data and saving the data. We make sure the Web Scraping System works properly. The Web Scraping System gets data from websites. This is a part of what the Web Scraping System does. The Web Scraping System is designed to get data from websites and save it.

3. Implementation Phase

Now we write the code for the Web Scraping System. We use Python to write the code. We also use tools like Requests and BeautifulSoup. We use Selenium and Pandas too. Python is a language for the Web Scraping System.

4. Testing Phase

We test the Web Scraping System to make sure it works correctly. The Web Scraping System needs to work, We test the Web Scraping System to make sure it gets all the information. The Web Scraping System is tested to ensure it works as expected. We test the Web Scraping System so we can trust the information it gives us. The Web Scraping System is important for getting information, about products.

Data Preprocessing and AnalysisData. Analysis

1. Collecting Information

The system gets product information from Flipkart pages. This includes the product name, the price, the ratings and the reviews of the product.

2. Data Cleaning

We clean the data to get rid of duplicates missing values and entries that do not make sense. This helps us do an analysis of the Data Preprocessing and Analysis.

3. Data Transformation

We change the data into a format that is easy to analyze. This is Data Transformation.

- **Text Processing is part of Data Transformation** : We process text data like product descriptions and reviews. We remove characters that we do not need from the Data Transformation.
- **Normalization is also important for Data Transformation** : We convert data into a format. This means we use numbers for the price and ratings of the product, in the Data Transformation.

Data Analysis

Descriptive Analysis : We examine the numbers to get to know our products, We are interested in details like product price. How ratings are distributed.

Correlation Analysis : We try to find out if there's a link between things. For instance does a products price affect its rating?

Feature Selection : We choose what to focus on, like product price, its rating and the number of reviews it has. We select these factors because they help us understand our products.

Visualization : We create charts and graphs to show how our products are performing and compare them. This helps us spot trends and patterns, in the data.

- ✓ **Bar charts are one way to do this.**
- ✓ **Word clouds are another way.**
- ✓ **Graphs also help to show the data patterns.**

5. Outcome

The system gets a lot of product information from Flipkart. It puts this information in an organized format. We can use the product information from Flipkart for things. For example we can look at the market compare prices of products and do research on products from Flipkart. The product data from Flipkart is really useful, for these things.

Proposed Algorithm

The way this system works is pretty simple. It does things in this order:

- ✓ **It sends a request to the Flipkart webpage.**
- ✓ **Then it gets the content of the HTML page.**

The system uses this content to do the steps. It uses BeautifulSoup to make sense of the HTML. After that it finds the product details that it needs from the HTML. It stores these details in a way that's easy to understand. The system does all these steps again for pages. Finally it cleans up the data it has collected and looks at it closely to learn something from the Flipkart product details.

Proposed Solution

The solution I am suggesting is a web scraping system built using Python. This system helps to extract product information from Flipkart.

- It collects product data in a way.

The collected data is then stored in a format. This structured format is useful for analysis of the product information, from Flipkart. The system works with Flipkart to extract product information.

Main Parts of the Web Scraping System

1. **Website Analysis** : I start by checking how Flipkart web pages are set up. I look for HTML tags and classes that have product info like product name, product price and product ratings.
2. **Data Extraction Module** : I extract product details.
3. **Data Storage Module** : I save the product data I extract in a format like CSV files or SQL databases so I can use it later and do more analysis on product data.
4. **Data Analysis Module** : In this part I look at the product data I've collected to find trends, patterns and insights.

For example I compare product prices. I see which products are popular on Flipkart. I check product prices. See how they rank on Flipkart. I analyze product ratings and product reviews.

Steps of Data Preprocessing

1. **Data Collection** : Collect product data from Flipkart using web scraping.
2. **Data Cleaning** : Remove incomplete or duplicate data entries.
3. **Data Transformation** : Convert raw data into structured formats.
4. **Data Integration**
5. **Feature Selection** : Select important features for analysis.

RESULT EVOLUTION AND ANALYSIS

The web scraping system we made can get product information from Flipkart pages. This information is put into organized lists that we can use to learn more about the products. The system makes things easier by doing the work of collecting data for us so we do not have to do it by hand. It also shows that web scraping can work well with sets of data, which is really useful, for the Flipkart web scraping system and the Flipkart product information.

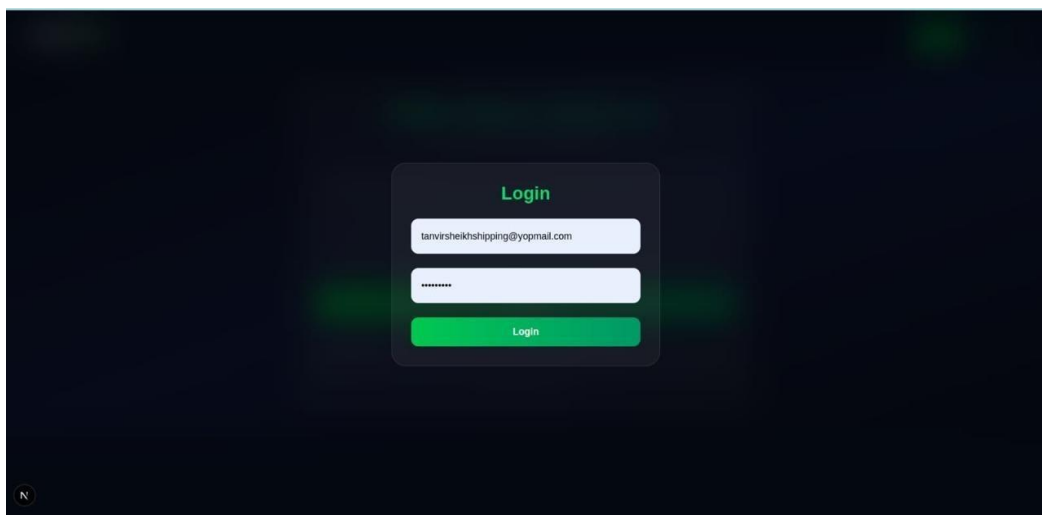


Figure 1: User Login Interface of the System

The picture shows the Login Interface of the web scraping system or Web Application. This is where people who have signed up can get into the system by putting in their email address and password. The login page looks really nice with a theme and a special box in the middle where you can log in. This page has spaces where you can put in your email and password and a button that says Login. When you click this button the system checks your information to make sure it is correct. If everything is okay you can then use the web scraping system and all its features, such as checking if newss real looking at it closely and using the dashboard.

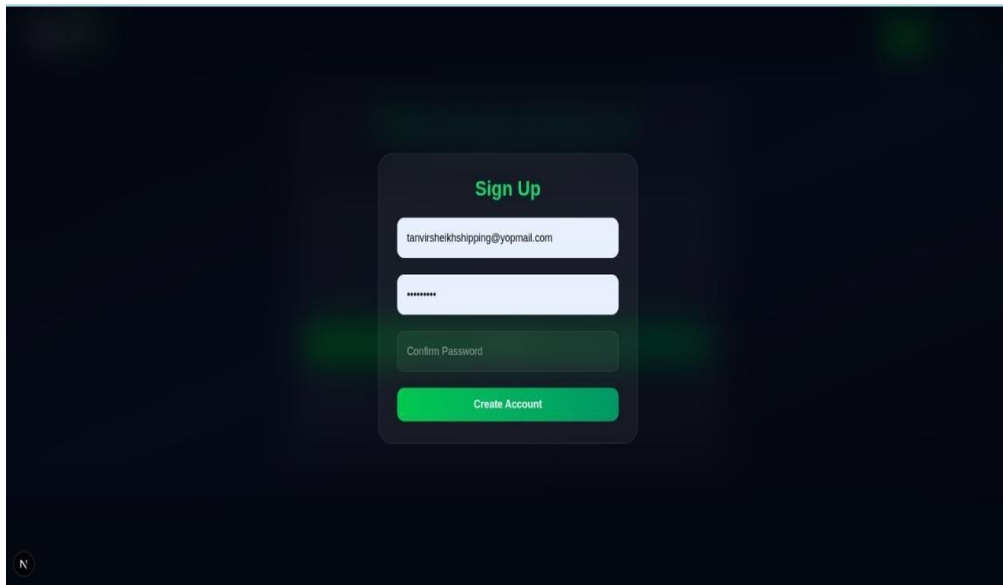


Figure 2: User Registration (Sign Up) Interface

The picture above shows how new users sign up to use the system. This is where people can create an account and get access to the system. The design of the system is easy to use. It looks clean. It has a background that makes the form easy to see on the system. When users are done filling in all their details they click the Create Account button to complete the sign up process for the system. Their details are then added to the system. After users have signed up for the system they can log in to the system using their email address and password for the system. The sign up part of the system is important for keeping user information safe and secure for all users of the system. It also makes sure that only the right people can use the system. The system keeps track of who's using the system and what they are doing on the system. Only people with an account can use the system. Only people with an account, for the system can use the system.

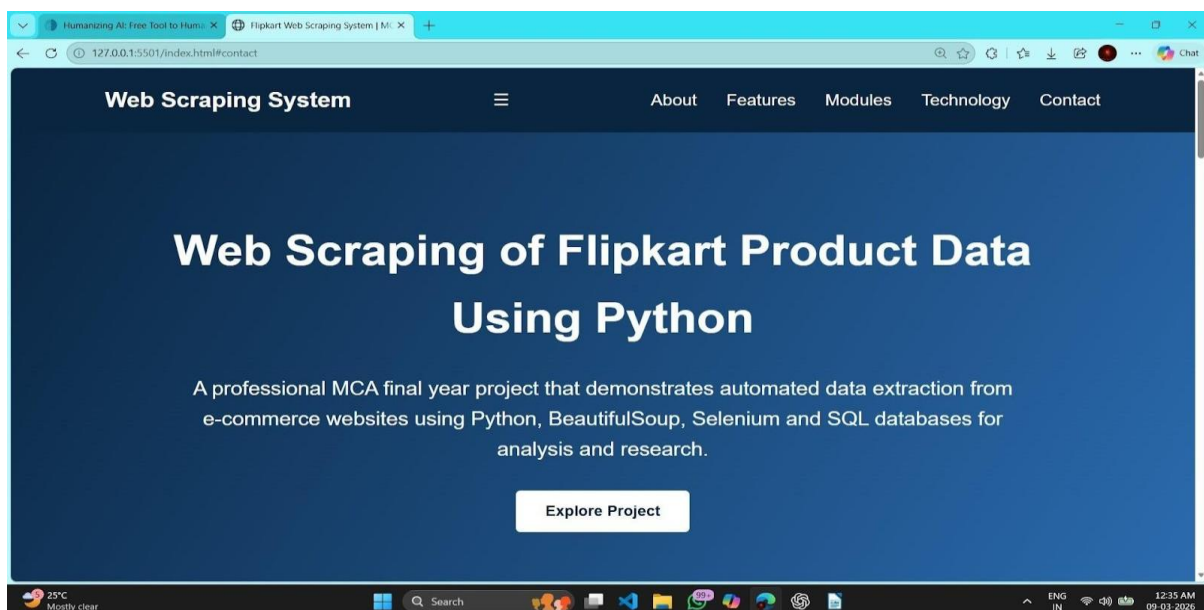


Figure 3: Main Analysis Interface of About the Web Scraping System

The Web Scraping System is a project that I worked on for my year of college. It is a system that can automatically get information from websites like Flipkart. This system uses a programming language called Python and some helpful tools like BeautifulSoup and Selenium to get details about products, such as what they're called how much they cost what people think of them and what people say about

them. The Web Scraping System then takes this information. Organizes it in a way that is easy to understand, like, in a big table or a special kind of file called a CSV file. This way people can look at the information later. Use it to make good decisions. The Web Scraping System is helpful because it saves people a lot of time and work. Normally someone would have to go to all these websites and write down all the information by hand.. The Web Scraping System can do it all by itself which is really cool. People can use the Web Scraping System for lots of things like comparing prices of products learning about what people like to buy and understanding what makes a product good or bad. The Web Scraping System is an example of how computers can help us make good choices by giving us the information we need.

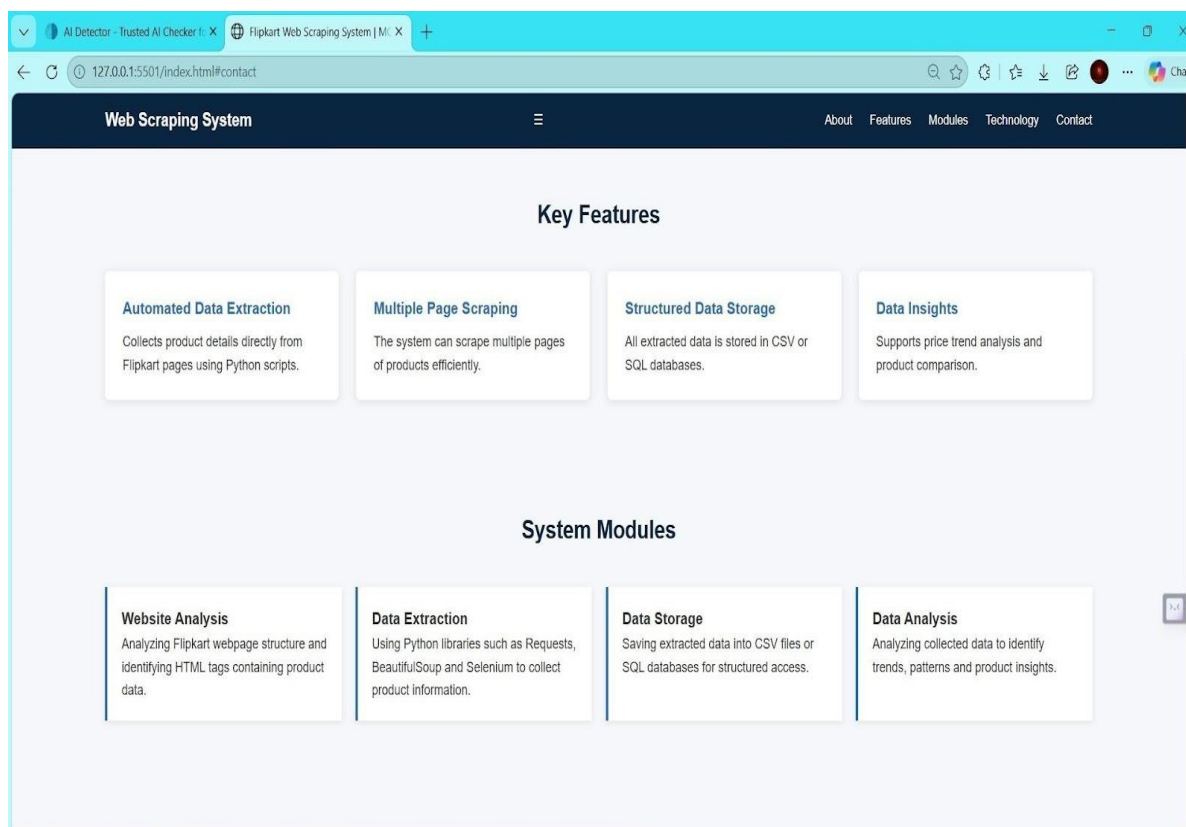


Figure 4: Key Features and System Modules of the Web Scraping System

Key Features

We use Python scripts to get product information from Flipkart. These scripts collect details like product name, price and rating. They also get reviews and descriptions. This way we do not have to enter data ourselves which makes the process more efficient. It also reduces mistakes. Our system can get data from product pages on Flipkart. It follows the links to these pages. Collects information. This helps us get a lot of product data quickly. We store all the product information we get in a format. We use files like CSV or databases like SQL. We can also analyze it easily. We look at the product data we collect to find information. We find things like price trends and what products are popular. We also see what customers like. This helps us compare products and understand what is happening in the market.

System Modules

In this part we look at the Flipkart website to see how it is organized. We find the HTML tags and CSS classes that have product details. We look for things like product name, price and rating. The data extraction part uses Python libraries to get webpage content. It uses libraries like Requests and BeautifulSoup and Selenium. It looks at the HTML code. Gets the product details we need. After we get the data we store it in files like CSV or databases like SQL. This part makes sure all the product

information is saved properly. We can then use it for analysis and other things. The data analysis part looks at the data we collect. It finds trends and patterns. We use it to look at product prices and ratings. We also look at customer reviews. This is useful, for market research and comparing products.

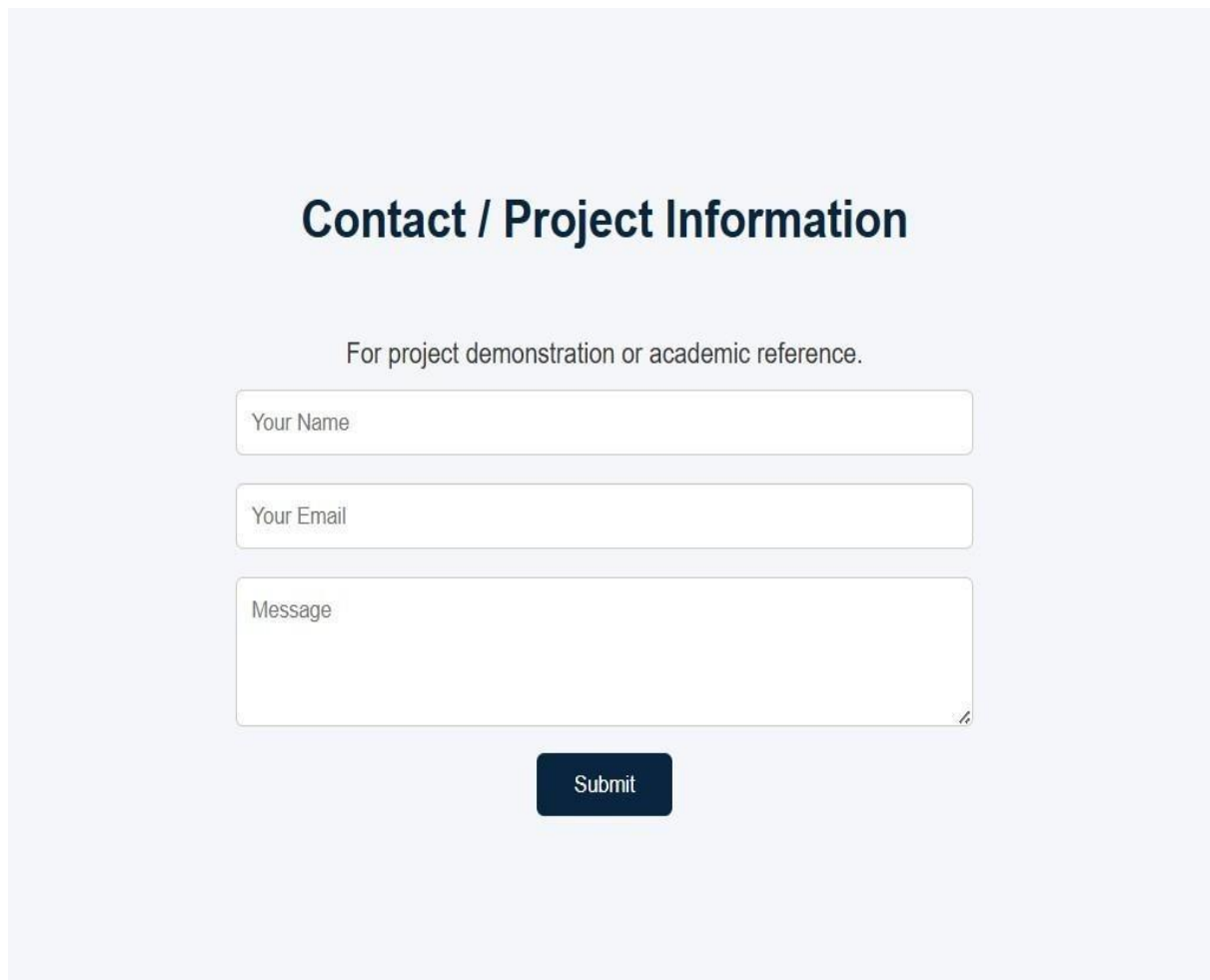
A screenshot of a web form titled "Contact / Project Information". The form is set against a light blue background. Below the title, there is a subtitle "For project demonstration or academic reference." followed by three input fields: "Your Name", "Your Email", and "Message". The "Message" field is a larger text area. At the bottom center of the form is a dark blue "Submit" button.

Figure 5: Contact us module of Web Scraping System

If you have any questions about the Web Scraping Product Data Using Python project please do not hesitate to contact us. We are happy to answer questions about how the Web Scraping of Flipkart Product Data Using Python project was done what technologies were used for the Web Scraping of Flipkart Product Data Using Python project or what we learned from the Web Scraping of Flipkart Product Data Using Python project.

The Web Scraping of Flipkart Product Data Using Python project is something we're very proud of. We look forward, to hearing from you. Talking more about the Web Scraping of Flipkart Product Data Using Python project.

Conclusion

The Web Scraping System for Flipkart Product Data using Python is really good at showing how automated technologies can be used to collect a lot of information from e-commerce websites like Flipkart. I think Flipkart is an example of an online marketplace. It generates a lot of product-related data every day including product names, prices, ratings, customer reviews and specifications from Flipkart. Collecting and organizing this data from Flipkart manually can be very difficult and time-consuming. The Web Scraping System for Flipkart Product Data using Python helps with this problem.

After preprocessing the cleaned dataset is stored in CSV files or SQL databases for Flipkart. This structured storage enables access, retrieval and further analysis of the collected information from Flipkart. The Web Scraping System for Flipkart Product Data using Python is also very useful for data analysis and insight generation of Flipkart data. The collected dataset can be used to analyze price variations across products on Flipkart compare product ratings on Flipkart and identify items based on customer reviews on Flipkart. The Web Scraping System for Flipkart Product Data using Python also provides experience in Python programming, database management and data processing for Flipkart. Students working on the Web Scraping System for Flipkart Product Data using Python gain hands-on knowledge about how real-world data from Flipkart's collected, processed and analyzed. It highlights the importance of automation and data-driven approaches in software systems and business analytics for Flipkart. However it is also important to consider aspects while performing web scraping on Flipkart. The Web Scraping System for Flipkart Product Data using Python focuses on extracting available information from Flipkart and respects the policies and guidelines of Flipkart. Responsible web scraping practices ensure that data collection does not harm Flipkart website performance or violate any terms of service of Flipkart.

In conclusion the Web Scraping System for Flipkart Product Data using Python successfully demonstrates how automated tools can be used to collect and analyze large-scale e-commerce data from Flipkart. The Web Scraping System for Flipkart Product Data using Python improves efficiency reduces work and provides a dataset that can be used for further research and analysis of Flipkart data. The Web Scraping System for Flipkart Product Data using Python serves as an example of how web technologies, data extraction techniques and data analysis methods can be combined to solve real-world problems related to Flipkart. For enhancements the Web Scraping System for Flipkart Product Data using Python can be extended by integrating features such as real-time data scraping from Flipkart, machine learning-based recommendation systems for Flipkart products, automated price tracking tools for Flipkart and interactive dashboards for data visualization of Flipkart data. These improvements can further increase the usefulness and scalability of the Web Scraping System for Flipkart Product Data using Python making it more powerful for research, business intelligence and data-driven decision-making, to Flipkart.

References:

1. R. Mitchell, *Web Scraping with Python: Collecting More Data from the Modern Web*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2018.
2. J. Lawson, *Web Scraping with Python: Collecting Data from the Modern Web*. Birmingham, U.K.: Packt Publishing, 2015.
3. C. Marrese-Taylor, M. Velásquez, and F. Bravo-Marquez, "A systematic study of web scraping techniques," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intelligence*, Warsaw, Poland, 2014, pp. 125–132.
4. G. L. Miller, "Data mining and web scraping for business intelligence," *International Journal of Computer Applications*, vol. 120, no. 5, pp. 1–6, 2015.
5. A. B. Mollah, M. A. Islam, and S. Hossain, "Automated web data extraction using Python for e-commerce analysis," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 4, pp. 381–387, 2020.
6. S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2016.
7. P. Chakrabarti, "Data mining for web data extraction and analysis," *ACM Computing Surveys*, vol. 45, no. 4, pp. 1–38, 2013.
8. F. Chollet, *Deep Learning with Python*. Shelter Island, NY, USA: Manning Publications, 2018.

9. A. Chaube, "ACO-Enhanced Siamese Networks for Robust Feature Matching in Copy-Move Image Forgery Detection," 2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA), Nagpur, India, 2024, pp. 1-6, doi: 10.1109/ICAIQSA64000.2024.10882433.