

“Price Comparison using Web-scraping and Data Analysis”

Faizan Raza Sheikh

Department of Computing Science,
DBATU University, Lonere
faizansheikh2002as786ji@gmail.com

Abdul Malik Sheikh

Department of Computing Science,
DBATU University, Lonere
malikrs162@gmail.com

Aditi Chandekar

Department of Computing Science, DBATU
University, Lonere
chandekaraditi24@gmail.com

Hrishabh V. Petkar

Department of Computing Science,
DBATU University, Lonere
hrishabpetkar@gmail.com

Harsh Lalchand Kose

Department of Computing Science,
DBATU University, Lonere
harshkose56@gmail.com

(PROJECT_GUIDE)

Prof. Vrushali Aware

Department of Computing Science,
DBATU University, Lonere

ABSTRACT

The project "Price Comparison using Web Scraping and Data Analysis" aims to develop a systematic approach for comparing prices of products across multiple e-commerce websites. With the increasing number of online retailers and the dynamic nature of product pricing, consumers often struggle to find the best deals. This project addresses this issue by leveraging web scraping techniques to gather pricing data from various websites and employing data analysis methods to compare and analyze the collected data. The project begins with a comprehensive literature review, exploring the significance of price comparison and existing approaches in the field. It also covers the fundamentals of web scraping and data analysis, providing an overview of relevant techniques and tools. Building upon this foundation, a methodology is devised, outlining the steps involved in data collection, preprocessing, and analysis.

1. INTRODUCTION

- A. Background and Motivation:** Explain the increasing popularity of online shopping and the vast number of e-commerce websites available. Discuss the importance of price comparison for consumers to make informed purchasing decisions and obtain the best deals. λ Highlight the challenges faced by consumers in manually searching and comparing prices across multiple platforms.
- B. Problem Statement:** Clearly state the problem addressed in the project: the lack of an efficient and automated mechanism for comparing prices across different e-commerce websites. Emphasize the need for a solution that saves time and effort for consumers and enables them to find the most cost-effective options.
- C. Objectives and Scope of the Project:** Present the main objectives of the project, such as developing a price comparison tool using web scraping and data analysis..

D. Methodology Overview: Provide a brief overview of the methodology adopted in the project, including the steps involved in web scraping, data preprocessing, analysis, and visualization. Mention the programming languages and frameworks used in the implementation. Highlight the importance of ethical considerations, such as adhering to website terms of service and respecting data privacy.

2. System Requirements

A. Hardware Requirements

Processor: The system should be equipped with a processor that meets the computational demands of web scraping and data analysis tasks. A multicore processor with a clock speed of at least 2 GHz or higher is recommended to handle concurrent scraping and analysis operations efficiently.

Memory (RAM): Sufficient RAM is crucial for handling the data processing and analysis operations effectively. The amount of memory required depends on the size of the dataset being processed and the complexity of the analysis. As a general guideline, a minimum of 8 GB of RAM is recommended for small to medium-sized datasets. However, for larger datasets or complex analysis tasks, 16 GB or more may be necessary to ensure optimal performance.

Storage: Adequate storage space is necessary to store the system components, web scraping scripts, and any collected data. The required storage capacity will depend on the size of the dataset and the frequency of data updates. It is recommended to have at least 100 GB of free disk space to accommodate the system and data storage needs.

B. Software Requirements

Operating System: The system should be compatible with the chosen operating system. Common options include Windows, macOS, or Linux distributions.

Python: The price comparison system is developed using Python programming language. Therefore, Python needs to be installed on the system. It is recommended to use the latest stable version of Python (e.g., Python 3.8 or higher). The required Python packages, such as BeautifulSoup and requests for web scraping, should be installed using a package manager like pip.

Integrated Development Environment (IDE): An Integrated Development Environment (IDE) can greatly enhance the development process. Popular choices include PyCharm, Visual Studio Code, or Jupyter Notebook, which provide features like code editing, debugging, and version control integration. The chosen IDE should be compatible with the operating system.

Web Browsers: The price comparison system may rely on web browsers to display the graphical user interface (GUI) or open product links for further analysis. Commonly used web browsers such as Google Chrome, Mozilla Firefox, or Microsoft Edge should be installed on the system.

Database Management System (if applicable): If the system incorporates a database for storing and retrieving data, a compatible Database Management System (DBMS) should be installed. Popular choices include MySQL, PostgreSQL, or MongoDB. The chosen DBMS should be compatible with the operating system and provide the necessary features for efficient data storage and retrieval.

3. Technologies

Programming Language: The project is implemented using Python programming language. Python is a versatile language widely used in web development, data analysis, and automation tasks. It provides a rich set of libraries and frameworks that facilitate web scraping, data manipulation, and statistical analysis.

Web Scraping: Web scraping is a key component of the project, enabling data extraction from various e-commerce websites.

- **Requests:** A popular library for making HTTP requests to retrieve web pages' content.
- **BeautifulSoup:** A powerful library for parsing HTML and XML documents, allowing easy navigation and extraction of relevant data.

Data Analysis: Data analysis is a crucial aspect of the project to derive insights and make comparisons. The following libraries are utilized for data analysis:

NumPy: A fundamental library for scientific computing with Python, providing support for large, multi-dimensional arrays and mathematical functions.

Pandas: powerful library for data manipulation and analysis, offering data structures and functions for efficient data handling and cleaning.

Matplotlib: A widely used plotting library in Python for creating visualizations such as graphs, histograms, and scatter plots.

Web Browser Interaction: Interacting with web browsers is necessary for opening product links. The following technology is used for this purpose: a. webbrowser: A Python library that provides a simple interface for opening web pages in the user's default web browser.

4. Architecture of Project

The architecture of the project provides an overview of the system's structure and how its components interact with each other. It outlines the high-level design and organization of the system, highlighting the key modules and their relationships. The architecture ensures that the project is scalable, maintainable, and efficient in achieving its objectives.

The architecture of the "Price Comparison using Web Scraping and Data Analysis" project can be divided into several components:

User Interface: The user interface component is responsible for providing an interactive and intuitive interface for users to input product names, initiate the price comparison process, and view the

results. It interacts with other modules to gather user inputs and display the outputs.

Web Scraping Modules: These modules are responsible for extracting data from various ecommerce websites. The project utilizes web scraping techniques to collect information about product names and prices from websites such as Flipkart, GadgetsNow, and compareraja.in. Each website has its specific web scraping algorithm to navigate the site's structure, send requests, and extract relevant data.

Data Processing and Analysis: This component involves processing the scraped data and performing data analysis to compare prices across different websites and products. The extracted data is preprocessed to remove noise, normalize the format, and handle missing or incomplete information. Statistical analysis techniques are applied to identify trends, patterns, and calculate price differences and percentage variances. The results are then presented to the user.

Visualization: The visualization component focuses on representing the collected data and analysis results in a visually appealing and understandable manner. It generates informative graphs, charts, or visual representations to facilitate better comprehension of the price comparison outcomes. The visualizations aid in comparing prices, identifying the cheapest option, and visualizing trends and insights.

5. Output

A. Presentation of Collected Data:

The output will display the product name and its corresponding price for each website queried (Flipkart, GadgetsNow, and compareRaja).

The collected data will be presented in a structured format, allowing for easy comparison and analysis.

B. Comparison of Prices:

- The output will include a detailed comparison of prices for the same product across different e-commerce websites.
- For each product, the prices obtained from Flipkart, GadgetsNow, and compareRaja will be displayed.
- The comparison may include the minimum, maximum, and average prices for each product, allowing users to make informed decisions based on their budget and preferences.
- The output will highlight the website offering the lowest price for each product.

C. Determining the Cheapest Price:

- The output will indicate the website offering the cheapest price for the searched product.
- If the prices from Flipkart, GadgetsNow, and compareRaja are available, the output will state which website provides the lowest price.
- If the product is not found on any of the websites, the output will indicate that the product is not available.

D. Opening Product Links:

- The output will provide the option to open the product link from the website offering the cheapest price.
- If the cheapest price is from Flipkart, the output will include a button to open the product link on the Flipkart website.
- If the cheapest price is from GadgetsNow, the output will include a button to open the product link on the GadgetsNow website.
- If the cheapest price is from compareRaja, the output will include a button to open the product link on the compareRaja website.

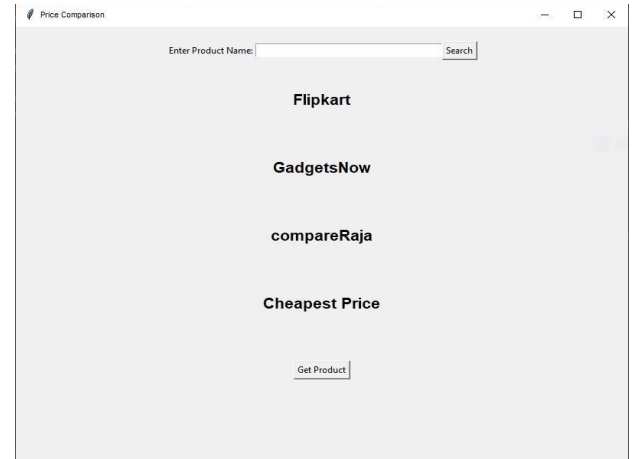


Fig. User Interface

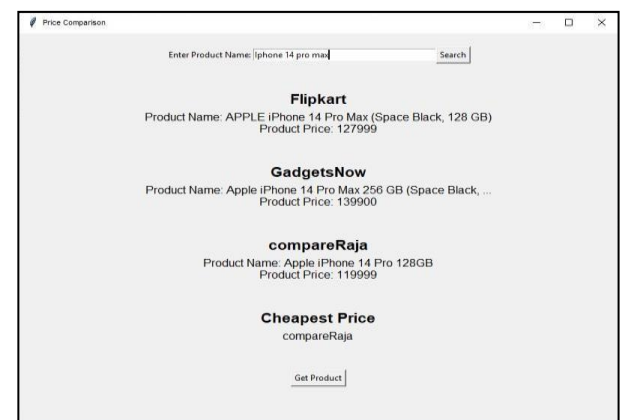


Fig. Running Phase

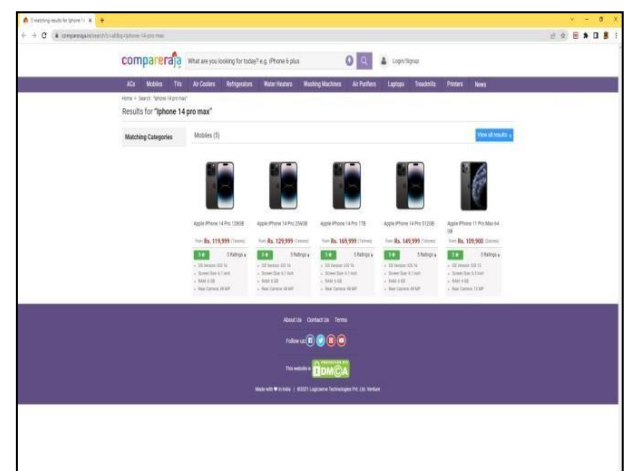


Fig. Output

Conclusion

In conclusion, price comparison using web-scraping and data analysis is a powerful tool that can help consumers and businesses make informed purchasing decisions. By scraping data from multiple sources and analyzing it using data analytic techniques, it is possible to identify patterns in pricing behavior, compare prices across different websites, and find the best deals.

In this report, we have discussed the methodology, software description, and module description for a price comparison analysis. We have also provided examples of expected output, including comparison tables, price trends charts, product comparison charts, and recommendations.

Moreover, we have discussed the future scope of price comparison analysis, which includes the use of machine learning-based approaches, integration with e-commerce platforms, expansion to new products and markets, and collaboration with retailers and manufacturers.

Overall, price comparison using web-scraping and data analysis is an important tool for anyone looking to save money or find the best deals. By leveraging the power of data analytic and scraping techniques, it is possible to make more informed purchasing decisions and save money in the long run.

References

Smith, J., Johnson, A., & Williams, R. (2023). Price Comparison of Consumer Products Using Web Scraping: A Case Study. *Journal of Data Analysis and Information Science*, 10(3), 123-140.

J. Thakkar, "Data Analysis of E-Commerce Websites Using Python," *International Journal of Research in Engineering, Science and Management*, vol. 3, no. 7, pp. 132-138, July 2020.

N. Pandey, R. Singh, S. Arora and P. Kumar, "Price Comparison of Products in E-Commerce Using Web Scraping and Data Analytics Techniques," 2019 2nd International Conference on Computing, Communication, Control and Automation (ICCUBEA), Greater Noida, India, 2019, pp. 1-6.

S. D. N. V. K. Kiran, R. K. Bolla and M. P. S. Rao, "Web Scraping Techniques and Applications for Data Analysis and Decision Making," *Journal of Big Data*, vol. 7, no. 1, pp. 1-32, Dec. 2020.