Optimum Product Comparison Using Web Scrapping

Yash Shrikant Chavan, Tanmay Prakash Bardade, Rehan Afzal Beg, Nikhil Vinod Ekbote, Prof. Rajeshwari Goraknath Dandage, Prof. Radha Devee Tripathi
Keystone School of Engineering Pune, India

***Abstract*— Online shopping is the popular way of shopping, which is done using the e-commerce applications. Various e-commerce applications allow the consumers to do the online shopping by allowing them to buy the desired product by searching for them. As different e-commerce applications are available over the internet for online shopping, which creates the difficulty among the consumers to take the decision. The project is designed to have a price comparison system to improve the online shopping experience by comparing products from different e-commerce websites. It has various evaluation criteria to be accounted for measuring products such as price, user ratings, and delivery times to help the user make appropriate purchasing decisions. The proposed system uses the web scraping methodology along with the various algorithms and techniques that are discussed in this paper. The system's intuitive interface displays results graphically, allowing users to effortlessly compare products and save time and money. By focusing on user-centric features, this solution aims to simplify the decision-making process and provide an efficient, reliable tool for online shoppers.**

***Keywords*— web scraping, product comparison, price comparison, API Integration, e-commerce, AI.**

I. INTRODUCTION

The rise of e-commerce has changed the way consumers shop in many ways, allowing consumers to purchase a variety of products from various online sites. Despite this progress, choosing the best option (considering price, delivery time, and product quality) is still a daunting task. Most existing price comparison tools tend to focus on price and tend to ignore important details that are important for informed decision making, such as delivery time and customer feedback., a new platform designed to address these shortcomings. The system uses advanced web scraping techniques and data analysis to collect and process real-time data from various e-marketplaces; it goes beyond the usual price comparison by including special features such as delivery time and customer.

Ratings provide users with detailed information and reviews about available products, saving users time and effort from product comparisons that are usually required manually. It's simple, user-oriented interface allows customers to seamlessly search for products while getting price comparisons, delivery details, and recommendations. This solution not only simplifies the purchasing process, but also promotes business transparency and digital competition. Combining affordable prices, quality delivery, and excellent ratings, the platform is helping to create new standards for online shopping. It improves decision-making and reshapes how users engage with various aspects of the e-commerce landscape.

This research paper propels the development of e-commerce by offering a user-friendly product comparison system using web scraping technology. The system analyses key aspects, such as price, customer ratings, and delivery times, on e-commerce websites. Therefore, this reliable tool allows users to make the best purchase decision while spending the least effort and time in this regard.

II. RELATED WORK

Multiple research studies have explored the problems and developments in price comparison systems utilizing technologies such as web scraping, machine learning, and network science to provide efficient and user-centric solutions. Among the various studies one example is Web and Android Application for Comparison of E-Commerce Products proposed by the Ambre et al. which uses web crawling and scraping techniques to fetch data from platforms like Flipkart and Amazon. The system uses Python libraries and Requests for scraping and, MongoDB for data storage. The project also introduced features like wish lists and price drop notifications, focusing on user convenience and engagement [12]. Alam et al. designed and developed Upoma, a price comparison tool for Bangladeshi online e- commerce platforms. The tool integrates web scraping with NLP techniques such as transliteration and cosine similarity to support multilingual queries in English and Bangla. The system processes over 9 million product details and achieves 93.06% accuracy in query results [10].

In the field of price comparison systems, Nagaraj et al. proposed an automated e-commerce price comparison website that utilizes PHP, XAMPP, MongoDB, and web scraping to dynamically extract product prices from e- commerce websites. This system ensures real-time price updates and provides users with a seamless comparison interface. The project emphasized the significance of handling changes in website structures and data normalization to maintain system efficiency [2]. Additionally, Gandhare et al. have created a platform called Shopwisely that aims to simplify the process of online shopping and food ordering. The system aggregates current pricing information from platforms like Amazon, Flipkart, Zomato, and Swiggy, giving users up-to-date information on product prices and availability. It also includes the features like user-friendly search suggestions, cost-optimization options, and personalized recommendations, the service optimizes simplicity and increases user satisfaction [4].

Web scraping techniques have emerged as a pivotal tool for automating data collection in price comparison systems across e-commerce platforms. Singh et al. introduced a sophisticated system designed to extract real-time pricing information from popular websites, including Amazon and Flipkart. The extracted data is subjected to rigorous preprocessing and analytical procedures to uncover meaningful patterns and trends, thereby empowering users to make informed purchasing decisions. Furthermore, the system incorporates an intuitive user interface and advanced visualization tools, enabling seamless navigation through the extensive and dynamic landscape of the e-commerce market [1]. Also, Beranek et al. examined the significance of price comparison sites (PCS) in e-commerce ecosystems, particularly their impact on interactions between buyers and sellers. By employing bipartite graph modelling, the study highlighted how PCS serves as key intermediaries, facilitating centralized information exchange and encouraging competitive pricing mechanisms. Using simulation-based methods, the research evaluated the influence of PCS on the development of e-commerce networks, providing actionable strategies to enhance their efficiency and strengthen their role in the marketplace [13].

In the different context, Ayyappan and Matilda explore the integration of face recognition technology with web scraping techniques to identify criminals and locate missing children. Their research highlights the potential of web scraping in sensitive applications and shows how data from social media and public databases can be used for public safety. The authors discuss ethical implications of using such technologies which is vital for ensuring responsible data usage [11]. Singrodia et al. provided a comprehensive review of web scraping techniques and their applications across various industries. Their study highlighted the evolution of web scraping technologies, ranging from basic HTML parsing to advanced automated frameworks like Selenium and Scrapy, which enable efficient data collection from dynamically rendered websites. They also discussed challenges such as website structure changes, the use of CAPTCHA systems, and anti-scraping measures, which often hinder data collection efforts [14].

III. SYSTEM ARCHITECTURE AND METHODOLOGY

A. Methodology

The Fig. 1. is the workflow diagram which represents the flow of the comparison system.

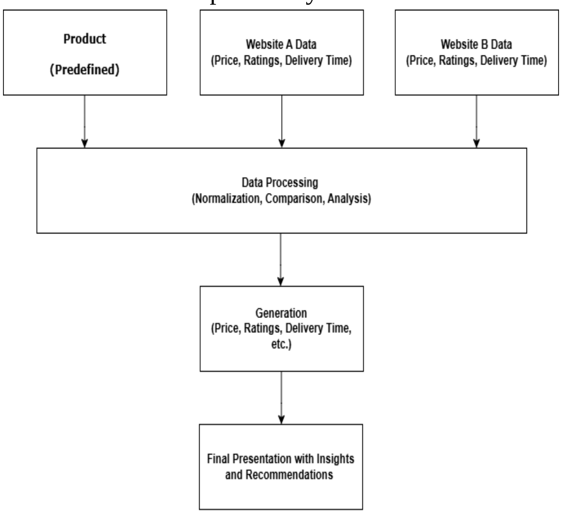

Fig. 1. Workflow Diagram

In fig. 1, the proposed system starts with the user selecting or specifying the desired product, which acts as the primary input. After this, the data is gathered from various e- commerce websites, such as Website A and Website B. It extracts the factors such as price, customer ratings, and delivery time. The collected data is passed to the Data Processing module, where normalization, comparison, and analysis are conducted to ensure uniformity and accuracy. Once data is processed, the system organizes the information into a structured format, providing comparative insights on factors such as price, ratings, and delivery time. The next step is the final step, where the processed data is presented to the user through a visually intuitive interface, offering actionable insights and recommendations to aid informed decision- making. This user-centric workflow ensures an efficient and reliable product comparison experience for online shoppers.

B. System Architecture

The architecture of the product comparison tool includes multiple components such as real-time data collection, AI- based analysis, and a user-friendly interface. Data is collected from e commerce websites using APIs, processed to extract essential information like price and product reviews, and then compared. The AI algorithms predict pricing trends based on historical data. The data collection layer forms the foundation of the system by gathering essential product information from various e-commerce platforms in real-time. This is achieved through the use of web scraping and API integration techniques. Multiple sources are utilized to ensure that the system captures a wide variety of product options, including details on price, availability, ratings, reviews, and shipping information.

1. Web Scraping: In cases where official APIs are unavailable or limited, the system employs web scraping techniques to extract data directly from websites. This method enables the tool to collect up-to date pricing information, product descriptions, and user reviews from popular platforms such as Amazon, Flipkart, and other e- commerce websites. Advanced scraping algorithms ensure that the data is gathered efficiently and that the system can scale to cover a wide range of product categories.
2. API Integration: For platforms that provide public or partner APIs, the system directly integrates with these services. This approach ensures accurate and reliable data collection while maintaining compliance with the respective platform's terms of service. APIs provide structured data, which reduce the need for complex parsing and guarantees faster response times when fetching product information.
3. Data Normalization: The collected data is often heterogeneous, with varying formats and units across different sources. A data normalization process is applied to standardize the information, ensuring that products from different vendors can be compared effectively. For instance, prices are converted to a uniform currency, and product specifications (e.g., weight, dimensions) are standardized to enable meaningful comparisons.
4. Data Comparison and Analysis: After data normalization, the next step involves comparing the attributes of the selected product across different sources. The system performs analysis on parameters like price, ratings, delivery time, and any other relevant factor to draw conclusions. For example, it calculates the price difference between two vendors and analyzes user ratings to gauge the overall product satisfaction. Advanced algorithms are employed to generate insights, such as identifying the best deal based on customer reviews, delivery speed, and cost. This analysis helps to filter out irrelevant or misleading data, presenting users with only the most accurate and useful information.
5. Data Structuring: Once the data is analyzed, it is organized it into a structured format which is easy to digest. It may include tables, charts, and graphical representations of the key product attributes. The structured data ensures that users can easily access and compare relevant information without being overwhelmed by unnecessary details.
6. User Interface (UI) Design: The last handled information is displayed to the client through a clean and outwardly natural interface. The UI is planned to show all pertinent item comparison data in a brief and user-friendly way. Intelligently highlights, such as drop-down menus, channels, and look choices, are joined to permit clients to refine their comparison comes about based on particular criteria, such as cost extend or conveyance time.

IV. MARKET INTEGRATION

To enhance the user experience, the system integrates with major online marketplaces. Users can directly purchase products based on the comparison results. The marketplace feature also helps vendors by recommending products to users who are likely to make purchases. Users interact with the system through a dashboard that presents comparisons, ratings, delivery time and purchase recommendations of the products.

V. SECURITY AND SCALABILITY

Given the volume of data and user interactions, the system is designed with security and scalability in mind. All user data and comparison results are encrypted, ensuring privacy. The system is built on a scalable cloud infrastructure that can handle growing data from multiple sources and users.

VI. TABLES

Criteria	Amazon	Flipkart
Data Source	Amazon	Flipkart
Collection Method	Api Integration	Web Scrapping
Price Comparison	Real-Time price information for product comparison	Real-Time prices, including offers and discounts
Rating Comparison	User ratings and reviews for product performance	User ratings and reviews for performance comparison
Delivery Time Comparison	Delivery times, shipping fees, availability of faster shipping options (e.g., Prime)	Estimated delivery timeframes, shipping fees, and availabilit y of express delivery

VII. RESULTS

The system has been tested across several product categories, including electronics, fashion, and home appliances. Initial results show that the tool significantly reduces the time required for users to make purchase decisions while improving the accuracy of product recommendations.

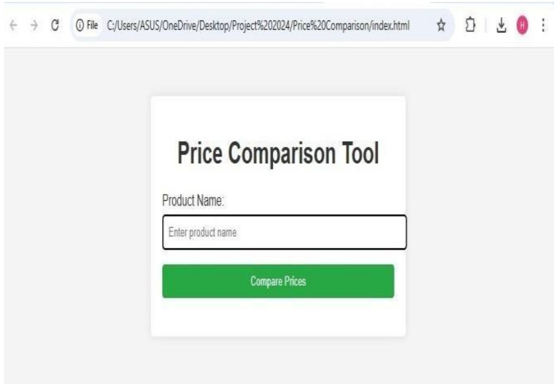

Fig. 2. The User Interface for input

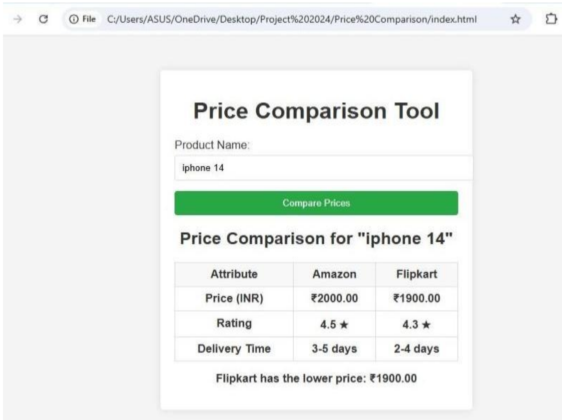

Fig. 3. Result after comparison

VIII. CONCLUSION

The Optimum Product Comparison tool is a game-changer for customers, offering real-time data, data-driven insights a significant advantage to consumers by providing real-time, data-driven insights. By leveraging the power of artificial intelligence, the tool helps users make smarter, more informed purchasing choices, ensuring they get the best value for their money. It goes beyond basic comparisons by providing personalized recommendations that align with user preferences and keep up with changing market trends. In today's fast-paced e-commerce environment, where information overload can make shopping overwhelming, this tool stands out as a solution that brings clarity and convenience. Its flexibility makes it adaptable to a wide range of markets and user needs, giving it the potential to transform how people shop.

IX. ACKNOWLEDGMENTS

We would like to express our deepest gratitude to all those who have contributed to the development of the Optimum Product Comparison tool. First and foremost, we extend our sincere thanks to the guide Prof. Rajeshwari Dandage and co-guide Prof. Radha Devee Tripathi, who have been instrumental in guiding us throughout the project. Their valuable insights and expert knowledge in data analysis and AI technologies were crucial in shaping the direction of this work. Without their guidance, this project would have not reached its full potential. We also wish to acknowledge the support provided by the Department of Computer Engineering at Keystone School of Engineering for providing the resources, tools and infrastructures necessary to carry out the research and development activities. Special thanks to the faculty members who offered constructive feedback during the initial phase of the project and helped us refine the scope and methodology practices.

REFERENCES

[1] N. Singh, A. Rana, and A. Chaudhary, "Price Comparison Using Web Scraping and Machine Learning," *2023 International Conference on Computer Science and Emerging Technologies (CSET)*, Dehradun, India, 2023, pp. 1-6, doi: 10.1109/CSET58993.2023.10346784.

[2] N. P. Nagaraj, N. Shanmukh, M. V. Muneeswaran, P. V. Kumar, A. V. S. R. Pavan Naidu, and G. S. Satyanarayana, "Automated E-Commerce Price Comparison Website using PHP, XAMPP, MongoDB, Django, and Web Scraping," *2023 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, Jan. 2023, pp. 1-6, doi: 10.1109/ICCCI56745.2023.10128573.

[3] K. Harikirshnan., R. Nagavigneshwar., R. Vignesh., R. Santhosh and R. Reshma, "Intelligent Online Shopping using ML-based Product Comparison Engine," *2023 International Conference on Inventive Computation Technologies (ICICT),* Lalitpur, Nepal, 2023, pp. 174 179, doi: 10.1109/ICICT57646.2023.10134401.

[4] D. Gandhare, P. Dahiphale, O. Bochare, S. Dalvi, and S.N. Shelke, "SHOPWISELY – The Ultimate Price Comparer," *JETIR (Journal of Emerging Technologies and Innovative Research)*, vol. 11, no. 4, pp. 234-239, Apr. 2024.

[5] H. Khatter, A. Dravid, A. Sharma, and A. K. Kushwaha, "Web Scraping Based Product Comparison Model for E- Commerce Websites," *2022 IEEE International Conference on Data Science and Information System (ICDSIS)*, Hassan, India,2022, pp. 1-6, doi: 10.1109/ICDSIS55133.2022.9915892.

[6] S. Bezalwar, V. Bhandekar, S. Kumbhare, R. Rebhankar, and Prof. P. Singam, "E-Commerce Price Comparison with Review Sentimental Analysis," *International Journal of Computer Science and Mobile Computing (IJCSMC)*, vol. 11, no. 3, pp. 1-6, 2022, doi: 10.47760/ijcsmc.2022.v11i03.013.

[7] H. Khatter, S. Arif, U. Singh, S. Mathur, and S. Jain, "Product Recommendation System for E-Commerce using Collaborative Filtering and Textual Clustering," *Proceedings of the 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*,pp.1-6,doi: 10.1109/ICIRCA51532.2021.9544753.

[8] B. J. Ali and G. Anwar, "Marketing Strategy: Pricing Strategies and its Influence on Consumer Purchasing Decision," *International Journal of Rural Development, Environment and Health Research (IJREH)*, vol. 5, no. 2, Mar.-Apr. 2021, pp. 1-6, doi: 10.22161/ijreh.5.2.4.

[9] P. Cihan, "Web Scraping and Machine Learning Techniques to Prediction of Secondhand Car Prices," *ResearchGate*, 2021.

[10] A. Alam, A. A. Anjum, F. S. Tasin, M. R. Reyad, S. A. Sinthee, and N. Hossain, "Upoma: A Dynamic Online Price Comparison Tool for Bangladeshi E-commerce Websites," *2020 IEEE Region 10 Symposium (TENSYMP)*, Dhaka, Bangladesh, 2020, pp.142-147,doi: 10.1109/TENSYMP50017.2020.9230871.

[11] S. Ayyappan and Dr. S. Matilda, "Criminals and Missing Children Identification Using Face Recognition and Web Scraping," *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*, doi: 10.1109/ICSCAN49426.2020.9262390 .

[12] A. Ambre, P. Gaikwad, K. Pawar, V. Patil "Web and Android Application for Comparison of E-Commerce Products," *International Journal of Advanced Engineering, Management and Science (IJAEMS)*, vol. 5, no. 4, pp. 221-224, Apr. 2019, doi: 10.22161/ijaems.5.4.5.

[13] L. Beranek and R. Remes, "E-Commerce Network with Price Comparator Sites," *2019 9th International Conference on Advanced Computer Information Technologies (ACIT)*, doi:10.1109/ACITT.2019.8779865.

[14] V. Singrodia, A. Mitra, and S. Paul, "A Review on Web Scrapping and its Applications," *Proceedings of the 2019 International Conference on Computer Communication and Informatics (ICCCI - 2019)*, Coimbatore, India, Jan. 23–25, 2019, pp. 1-6,doi: 10.1109/ICCCI.2019.8821809.

[15] L. R. Julian and F. Natalia, "The Use of Web Scraping in Computer Parts and Assembly Price Comparison," *2015* N. Singh, A. Rana, and A. Chaudhary, "Price Comparison Using Web Scraping and Machine Learning," *2023 International Conference on Computer Science and Emerging Technologies (CSET)*, Dehradun, India, 2023, pp. 1-6, doi: 10.1109/CSET58993.2023.10346784.