iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59194

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Web Scraping Application

Nitya Sharma[1], Nishant Agnihotri[2]
BBDNIIT, Lucknow

Abstract: ***The web scraping application is designed to empower users in tracking and monitoring the prices of products on the Amazon platform. Utilizing advanced web scraping techniques with Cheerio and Axios, the application allows users to add specific products for real-time price monitoring. Leveraging the Next.js framework for efficient web development, the application integrates MongoDB for seamless data storage. To overcome potential access restrictions, the application employs proxy services provided by Bright Data. The inclusion of Cron Jobs facilitates automation and scheduling, ensuring timely updates of product prices. For user engagement and communication, Node Mailer is utilized to send email notifications about price changes. Additionally, the application offers features such as historical price data visualization and user preferences management. Through a combination of these tools and technologies, application aims to provide users with a comprehensive and user-friendly solution for effective price tracking and informed decision-making on the Amazon platform.***

I. INTRODUCTION

The project represents an innovative web scraping application designed to address the growing demand for efficient and real-time monitoring of product prices on the Amazon platform. In today's dynamic e-commerce landscape, where prices can fluctuate rapidly, staying informed about product pricing is crucial for both consumers and businesses. Leveraging cutting-edge technologies such as Cheerio and Axios for web scraping and Next.js for web development, our application empowers users to add specific products for continuous, real-time price tracking. MongoDB ensures seamless data storage, while Bright Data's proxy services overcome potential access challenges. The integration of Cron Jobs automates the process, providing users with timely updates. Additionally, Node Mailer enables personalized email notifications, enhancing user engagement. With features including historical price data visualization and user preferences management, this project aims to offer a robust solution for users seeking accurate and timely insights into product pricing dynamics on Amazon.

II. PROBLEM STATEMENT

The e-commerce landscape, particularly on platforms like Amazon, is characterized by dynamic and frequently changing product prices. The absence of an efficient and user-friendly tool for real-time monitoring poses a significant challenge for consumers and businesses alike. Current solutions often lack the versatility needed to adapt to Amazon's evolving website structure and frequently encounter issues related to access restrictions. Furthermore, the absence of effective notification mechanisms hampers users' ability to promptly respond to price fluctuations. This project addresses these challenges by developing the application, a web scraping tool equipped with advanced technologies to provide users with accurate, real-time updates on product prices, overcoming common barriers associated with web scraping and ensuring a seamless user experience.

III. JUSTIFICATION

The development of the application is justified by the clear need for an efficient and reliable tool to address the challenges inherent in monitoring and tracking product prices on the dynamic Amazon platform. Existing solutions, while offering certain features, often fall short in providing real-time updates, adapting to changes in website structure, and ensuring a seamless user experience. The advanced technologies employed, including Cheerio and Axios for web scraping, Next.js for web development, and MongoDB for data storage, signify a commitment to creating a robust and versatile solution. The integration of proxy services from Bright Data and automation through Cron Jobs further enhances the application's capability to overcome access restrictions and ensure timely updates. By incorporating user-friendly features such as personalized email notifications and historical price data visualization, application aims to not only fulfil the current gaps in the market but also set a new standard for effective, ethical, and user-centric e-commerce price monitoring applications.

The justification lies in providing users with a powerful, accessible, and reliable tool to navigate the complexities of online shopping and make informed purchasing decisions.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

IV. RELATED WORK

In the realm of web scraping applications and e-commerce price monitoring, several existing solutions have attempted to address the challenge of tracking dynamic product prices on platforms like Amazon. Tools have gained popularity for their ability to provide historical price trends and offer browser extensions for price tracking. However, these solutions often face limitations in adapting to changes in Amazon's website structure and may not provide real-time updates. Additionally, challenges related to access restrictions and the potential violation of website terms of service raise concerns about the long-term viability and ethical considerations of such tools. While existing solutions contribute valuable features, the project aims to go beyond by integrating advanced web scraping technologies, proxy services, and automation features to offer a comprehensive and user-friendly application that addresses the limitations observed in current tools. The objective is to create a robust platform that not only overcomes technical challenges but also prioritizes user experience and ethical web scraping practices.

V. FUNCTIONAL REQUIREMENTS

The application ensures a user-friendly experience through key functionalities. It provides secure user authentication, real-time updates on monitored product prices, and automated email notifications for significant price changes. The system also stores historical price data, facilitating trend analysis. The intuitive user interface enhances navigation, and automation, including periodic updates, ensures data accuracy. Upholding data privacy standards, the application adheres to regulations while delivering a streamlined solution for real-time price monitoring on Amazon.

1) *User Authentication:* Establish a secure foundation with a seamless account creation and login process, ensuring user data integrity and protection.
2) *Real-Time Price Updates:* Provide users with instant, accurate, and real-time updates on the pricing dynamics of their monitored products on Amazon.
3) *Email Notifications:* Enhance user engagement through automated email alerts, promptly notifying users of noteworthy changes in product prices.
4) *Historical Price Data:* Enable users to gain valuable insights by storing and presenting historical price data, empowering informed decision-making based on past trends.
5) *User Interface (UI):* Foster a positive user experience with an intuitive and user-friendly interface, allowing seamless navigation and interaction.
6) *Automation:* Streamline user experience by implementing automated processes, including periodic updates through Cron Jobs, reducing manual intervention.
7) *Data Privacy:* Uphold user trust and comply with privacy standards and regulations, ensuring the secure handling and protection of user data throughout the application.

VI. NON-FUNCTIONAL REQUIREMENTS

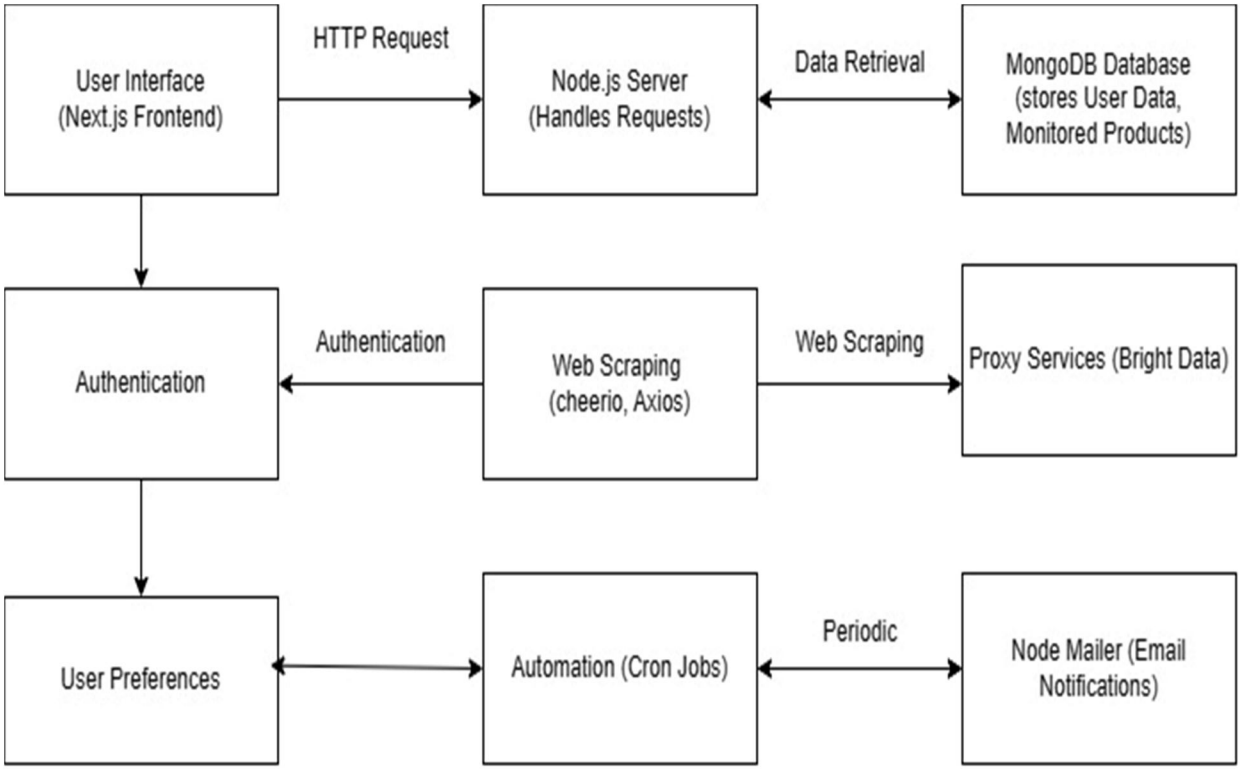
1) *Performance:* Ensure the application delivers real-time updates with minimal latency, guaranteeing a seamless and responsive user experience even during peak usage.
2) *Scalability:* Design the system to effortlessly scale, accommodating a growing user base and expanding product monitoring without compromising performance.
3) *Reliability:* Maintain a reliable system with minimal downtime, providing consistent availability to users relying on up-to-date pricing information.
4) *Usability:* Prioritize an intuitive and user-friendly interface, facilitating easy navigation for users of varying technical proficiency.
5) *Security:* Implement robust security measures to safeguard user data, ensuring privacy and preventing unauthorized access.
6) *Availability:* Guarantee 24/7 availability, allowing users uninterrupted access to real-time price updates whenever they need.
7) *Maintainability:* Design the application with maintainability in mind, enabling efficient updates, bug fixes, and future enhancements.
8) *Portability:* Ensure the application's portability across diverse devices and platforms, enhancing accessibility for a broad user base.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

VII. SYSTEM DESIGN

The system is architecturally designed to provide users with a seamless and efficient experience in monitoring and tracking product prices on the Amazon platform. The frontend, developed using the Next.js framework, ensures a robust user interface with dedicated pages for tasks such as product monitoring, historical data visualization, user preferences, and notifications. On the backend, Node.js handles user interactions and requests, while MongoDB serves as the database for storing user data, monitored products, and historical price records. The web scraping functionality relies on Cheerio for parsing HTML and Axios for fetching real-time price information from Amazon, with Bright Data's proxy services integrated to overcome potential access restrictions. Automation is facilitated by Cron Jobs, ensuring timely updates of product prices without manual intervention. Node Mailer handles email notifications to users, enhancing communication about significant price changes. User preferences and authentication are secured within the system, and historical price data is visualized using interactive charts. Robust error handling, logging mechanisms, and adherence to ethical web scraping practices contribute to a secure and reliable system. The user interface is designed for intuitive navigation and a responsive experience, aiming to provide a comprehensive and user-centric solution for efficient e-commerce price monitoring on Amazon.



VIII. CONCLUSION

The project presents a promising solution to the challenges associated with real-time monitoring of product prices on the Amazon platform. By leveraging advanced web scraping technologies like Cheerio and Axios, incorporating Next.js for efficient web development, and integrating MongoDB for seamless data storage, the application stands poised to offer users a comprehensive and user-friendly experience. The utilization of Bright Data's proxy services addresses potential access restrictions, while the inclusion of Cron Jobs ensures automated and timely updates of product prices. Node Mailer enhances user engagement by providing personalized email notifications. Through features such as historical price data visualization and user preferences management, application aspires to become a go-to tool for users seeking accurate and timely insights into dynamic pricing dynamics on Amazon. As we move forward, continuous refinement and adherence to ethical web scraping practices will be pivotal to the project's success, ensuring it remains both effective and respectful of the broader online ecosystem.

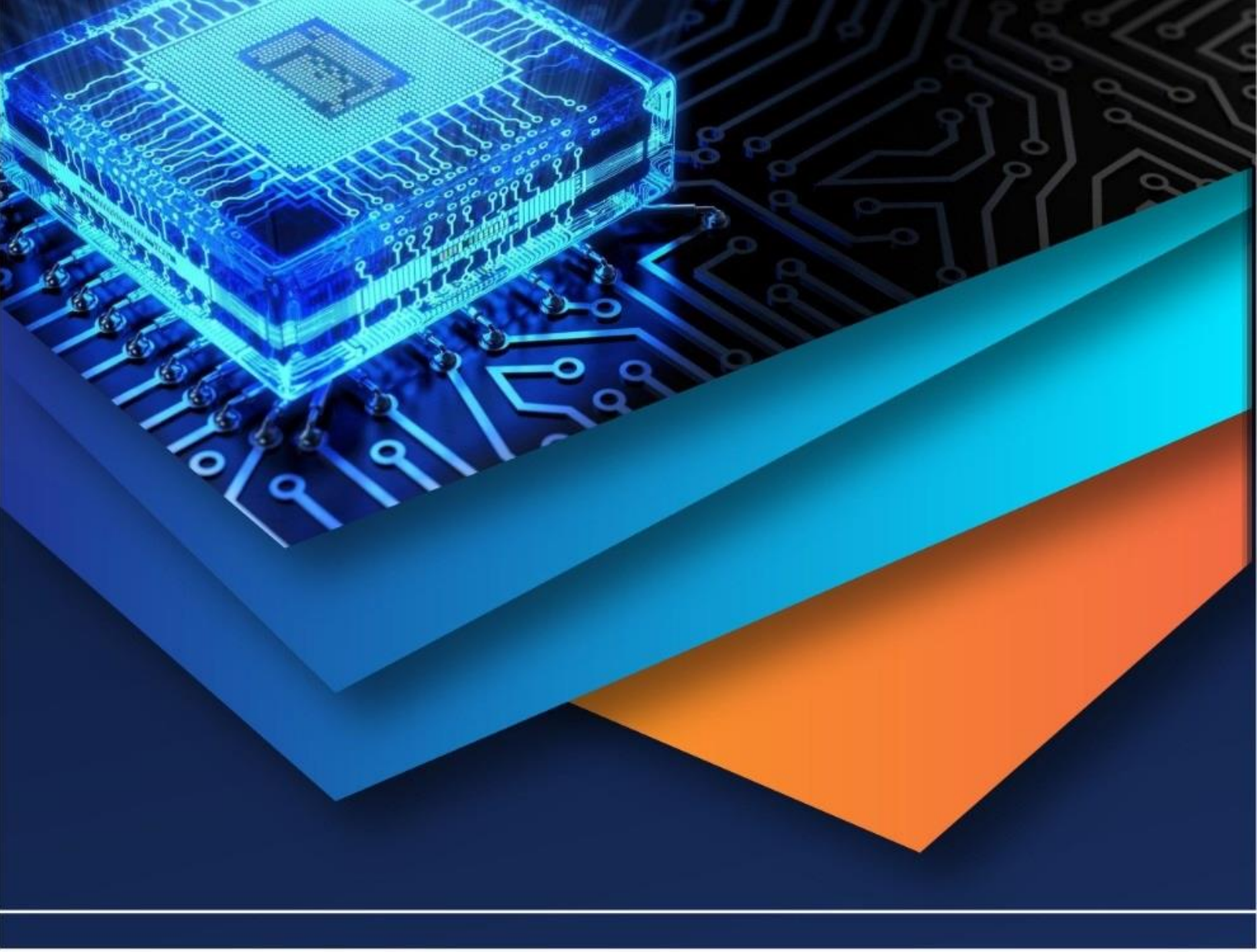
The application amalgamates cutting-edge functionality with user-centric design, offering a dynamic and reliable solution for real-time price monitoring on Amazon. Through robust features and meticulous attention to non-functional requirements, this application stands poised to redefine the user experience in the realm of e-commerce.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] HTML: W3Schools. (2023). HTML Tutorial
[2] CSS: MDN Web Docs. (2023). CSS Reference
[3] JavaScript: JavaScript.info. (2023). The Modern JavaScript Tutorial
[4] Next.js: Next.js Documentation. (2023). Getting Started
[5] MongoDB: MongoDB Documentation. (2023). MongoDB Manual
[6] Proxy Services: Bright Data. (2023). Bright Data Documentation
[7] Web Scraping:
[8] Cheerio: Cheerio Documentation
[9] Axios: Axios Documentation
[10] Automation and Scheduling: Cron Jobs. (2023). Cron Jobs Documentation
[11] Email Notification: Node Mailer. (2023). Node Mailer Documentation
[12] Favicon: Favicon Generator . Favicon Generator

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)