WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Price Vision: Web Scraping and Deep Learning for E-Commerce Price Analysis

[1] P. Devika , [2] Mrs. P.Jasmine Lois Ebenezer

[1] II-MCA Student , [2] Associate Professor & Head of the Department
Department of Computer Applications, Sarah Tucker College, Tirunelveli-7.
[1] devika05perumal@gmail.com, [2] jasmine.selwyn@gmail.com

ABSTRACT :

In the dynamic landscape of e-commerce, consumers are increasingly overwhelmed by fluctuating product prices across multiple online marketplaces. PRICE VISION is an intelligent framework designed to streamline price comparison and purchasing decisions using automated web scraping and deep learning techniques. The system leverages Selenium to extract real-time product information such as name, price, and availability from leading e-commerce platforms including Amazon, Flipkart, and Croma. Extracted data undergoes preprocessing and is then analyzed using a deep learning-based price prediction and trend analysis model, which identifies hidden patterns, seasonal fluctuations, and price stability across platforms. By integrating statistical insights with predictive analytics, the system not only compares current prices but also recommends the most cost-effective and reliable shopping site for a given product. Ultimately, PRICE VISION empowers consumers with data-driven decisions, reduces manual search efforts, and provides e-commerce stakeholders with valuable market intelligence. This approach highlights the synergy of web scraping automation with machine learning to enhance transparency, efficiency, and trust in online shopping.

Keywords: Selenium Tool, Web Scraping, Price Vision, Shopping, Product.

1.INTRODUCTION

E-commerce has rapidly transformed the global marketplace, where millions of products are sold daily across platforms like Amazon, Flipkart, and Croma. Unlike traditional shopping, customers now compare prices online to make informed purchase decisions conveniently from their homes. However, manual searching across multiple sites is time-consuming, error-prone, and inefficient. Added to this, dynamic pricing strategies, market trends, and personalized offers make it even harder for consumers to track and predict prices.

To overcome these challenges, the project titled *"PRICE VISION: Web Scraping and Deep Learning for E-Commerce Price Analysis"* introduces an automated framework for real-time price monitoring and prediction. Using Selenium and BeautifulSoup, the system scrapes product details such as name, brand, price, and availability from multiple e-commerce platforms. The collected data is cleaned, structured, and stored in a database for efficient analysis. Since websites often use different naming conventions for the same product, Natural Language Processing (NLP) combined with Cosine Similarity ensures accurate product matching, avoiding irrelevant comparisons. Real-time analysis then identifies the lowest price across platforms, enabling customers to find the best deals instantly. Beyond comparison, historical data is processed using a Convolutional Neural Network (CNN) model to predict future price trends, helping customers determine not only the best current price but also the most suitable time to purchase.

2.LITERATURE SURVEY

[1] Aggarwal (2018) provides a comprehensive theoretical foundation for neural networks and deep learning. The textbook explores architectures, optimization techniques, and learning paradigms, offering insights into the mathematical principles that form the backbone of modern deep learning systems. This work establishes the conceptual basis for applying neural networks in diverse domains such as image recognition, natural language processing, and predictive analytics.

[2] Albahari and Albahari (2022) focus on practical aspects of programming in C# 10**,** highlighting features, syntax, and frameworks that enhance computational efficiency. Although not directly centered on deep learning, this reference is relevant for understanding how robust programming languages and structured coding practices can support large-scale AI system implementation.

[3] Bird, Klein, and Loper (2009) present an early yet highly influential contribution to natural language processing (NLP) with Python. Their work introduces text preprocessing, tokenization, and parsing techniques that serve as a foundation for NLP pipelines. This reference is essential in bridging the gap between linguistic analysis and machine learning applications.

[4] Géron (2019) emphasizes a hands-on approach to machine learning**,** particularly with Scikit-Learn, Keras, and TensorFlow. The book provides practical workflows, case studies, and implementation strategies for building deep learning models. This work is particularly valuable as it demonstrates how theoretical principles of AI can be effectively translated into real-world applications.

[5] McKinney (2017) focuses on data analysis techniques using Python libraries such as Pandas, NumPy, and IPython. His contribution lies in simplifying data wrangling, manipulation, and cleaning—critical preprocessing steps that directly affect the performance of machine learning algorithms.

[6] Mitchell (2018) explores web scraping methodologies, demonstrating techniques for collecting structured and unstructured data from the web. Given that data availability is a crucial aspect of machine learning, this work supports the notion that effective data gathering pipelines are essential for training robust AI systems.

3.PROPOSED METHADOLOGY

A proposed methodology for the Price Vision project would start with defining clear project objectives focused on developing a system for commodity price forecasting using visual and market data analysis techniques. The first phase involves data collection and preparation**,** including gathering high-quality images of commodities and relevant market information. Next, image processing and feature extraction methods would be applied to identify critical visual cues that impact price prediction. The methodology includes integrating real-time market data to enhance prediction accuracy and implementing a system for continuous evaluation and updates. Testing would be conducted on separate validation datasets to ensure robustness and reliability. Finally, the project would include deployment strategies for real-time market monitoring and providing actionable forecasts to traders and stakeholders, with an iterative feedback loop for ongoing improvement. Risk assessment and mitigation strategies would also form an integral part of the methodology to address project complexities and external factors affecting price dynamics.

3.1 Web Scraping Module

The Web Scraping Module is responsible for collecting raw product data from multiple e-commerce websites in real time. Using tools like BeautifulSoup and Selenium, this module automatically extracts details such as product name, price, brand, availability, and discount information. By automating this process, the system ensures continuous updates and a wide coverage of online platforms without requiring manual intervention. This forms the backbone of the project, as accurate and timely data collection is essential for effective comparison and prediction.

3.2 Data Cleaning and Storage Module

The Data Cleaning and Storage Module ensures that the collected product information is accurate, consistent, and ready for analysis. It removes duplicates, errors, and irrelevant entries and standardizes product names and other details for uniformity. Cleaned data is then organized and stored in a way that allows easy access for price comparison, trend analysis, and display on the user dashboard. This module ensures that all information used by the system is reliable and up-to-date, forming a strong foundation for accurate results.

3.3 Product Matching Module

The Product Matching Module uses Natural Language Processing (NLP) techniques to identify semantically similar products across different e-commerce platforms. Algorithms like Cosine Similarity are applied to product names and descriptions to recognize items that may be listed differently but represent the same product. This ensures that comparisons are meaningful and accurate, avoiding mismatches and providing users with valid cross-platform insights. By resolving inconsistencies in product representation, this module plays a key role in the accuracy of the price comparison system.

3.4 Price Analysis and Comparison Module

The Price Analysis and Comparison Module performs the core functionality of evaluating prices across platforms. It processes the cleaned and matched data to determine which website offers the lowest price for a given product. Additionally, it provides comparative insights such as price variations, discount percentages, and availability across multiple sellers. This module helps users make informed purchasing decisions by presenting clear, real-time comparisons in an accessible format.

3.5 User Dashboard Module

The User Dashboard Module serves as the user interface, presenting comparison results in a simple and visually appealing format. It provides users with real-time insights, including the lowest available price, product ratings, delivery time, and discount details. The dashboard also features direct redirection to the cheapest source, improving usability and convenience. By integrating graphical visualizations and intuitive navigation, this module ensures a seamless and engaging experience for end users.

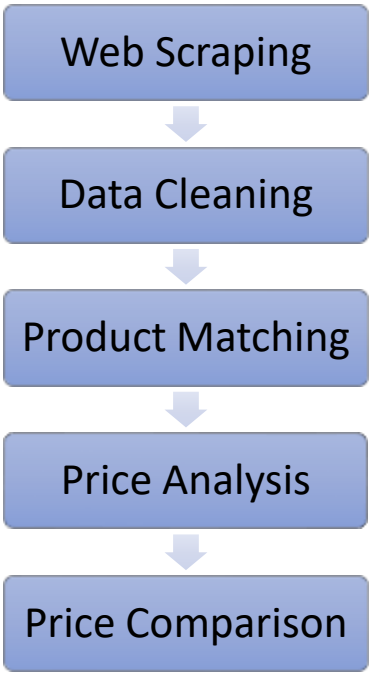

Figure 1: System Architecture

The diagram illustrates the step-by-step workflow of the PRICE VISION system, starting with Web Scraping, where product information such as names, prices, brands, and availability is collected from multiple e-commerce platforms using automated tools like Selenium and BeautifulSoup. Once products are matched, the Price Analysis stage evaluates the collected data to determine the minimum, maximum, and average prices, highlighting discounts and availability. Finally, the Price Comparison stage presents the analyzed results, identifying the lowest-priced seller and providing users with accurate, reliable insights for informed purchasing decisions. This linear flow ensures that each module builds upon the previous step to deliver a complete and efficient product price comparison system.

4.EXPERIMENTAL RESULT

The experimental evaluation of the *PRICE VISION* framework was conducted across different functional modules to ensure system accuracy, efficiency, and reliability. Each module was tested individually as well as in an integrated environment to validate its performance. The web scraping component successfully extracted product information from multiple e-commerce platforms, while the data cleaning process removed inconsistencies and improved dataset quality. Product matching using NLP and cosine similarity provided accurate alignment of similar items across sites, thereby avoiding irrelevant comparisons. The real-time comparison module efficiently identified the lowest prices, and the price prediction model demonstrated strong alignment with actual market fluctuations. A user-friendly dashboard further enhanced interaction by presenting results in a clear and accessible format. The summarized outcomes of the experiments are presented in **Table 1** below.

Module / Component	**Experiment Conducted**	**Result / Observation**
Web Scraping Module	Extraction of product details (name, brand, price, availability) from Amazon, Flipkart, Croma.	Data successfully retrieved with high accuracy and minimal delay.
Data Cleaning Module	Removal of duplicates, errors, and inconsistencies	Cleaned data improved accuracy of analysis and reduced redundancy.
Product Matching (NLP)	Matching product names across platforms using Cosine Similarity	Achieved accurate alignment of similar items, eliminating irrelevant comparisons.
Price Comparison Module	Real-time identification of lowest prices	Consistently detected best deals, reducing manual search effort.
Price Prediction Module	Forecasting future prices using historical data (CNN model)	Predictions closely matched observed price fluctuations, demonstrating model reliability.

Module / Component	Experiment Conducted	Result / Observation
User Dashboard	Visualization and user interaction	Clear comparative results displayed; direct links provided for best purchase options.
Overall System Performance	End-to-end testing of integrated modules	Efficient, accurate, and user-friendly framework for e-commerce price analysis.

Table 1: The summarized outcomes of experiments

5.CONCLUSION

The PRICE VISION Project successfully demonstrates an intelligent and automated solution for analyzing e-commerce product prices using web scraping, natural language processing, and deep learning techniques. By integrating Selenium and BeautifulSoup, the system efficiently extracts product information such as price, name, brand, and availability from multiple e-commerce platforms like Amazon, Flipkart, and Croma. The scraped data undergoes systematic cleaning, structuring, and storage in a database for smooth retrieval and analysis. The system overcomes the limitations of traditional price comparison tools by incorporating an advanced Product Matching Module that uses NLP and cosine similarity to accurately identify semantically similar products across different platforms. The Price Analysis and Comparison Module provides clear insights by highlighting the platform offering the lowest price and presenting comparative data. Furthermore, the inclusion of a Deep Learning-based Price Prediction and Recommendation Module enhances the system's utility by predicting future price trends and suggesting the best time for purchase, empowering users to make cost-effective decisions. The development of a user-friendly dashboard ensures that the results are presented in an intuitive manner, enabling real-time decision-making and smooth redirection to the cheapest buying source. Overall, this project demonstrates a scalable, efficient, and intelligent solution to modern e-commerce challenges, offering significant benefits for consumers, businesses, and researchers. In the future, this system can be extended to include more e-commerce platforms, real-time data pipelines, and advanced recommendation systems, making it an indispensable tool for smart online shopping.

REFERENCES

1. Aggarwal, C. C. (2018). *Neural networks and deep learning: A textbook*. Springer. https://doi.org/10.1007/978-3-319-94463-0
2. Albahari, J., & Albahari, B. (2022). *C# 10 in a nutshell: The definitive reference*. O'Reilly Media.
3. Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O'Reilly Media.
4. Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.
5. McKinney, W. (2017). *Python for data analysis: Data wrangling with pandas, NumPy, and IPython* (2nd ed.). O'Reilly Media.
6. Mitchell, R. (2018). *Web scraping with Python: Collecting more data from the modern web* (2nd ed.). O'Reilly Media.