

**International Journal of Research Publication and Reviews** 

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

# **Online Ecommerce Product Price Comparison Via Meachine Learning**

## Rokesh.D

Master Of Computer Application, M.G.R Educational And Research Institute , Chennai, Tamilnadu, Email : <u>davidrokesh@gmail.com</u>

#### Abstract

This project presents a simple yet effective price comparison tool for online e-commerce products using Google Colab. The system enables users to input a product name and compare prices across multiple e-commerce platforms by utilizing pre-collected or lightly scraped data. The data is processed and cleaned in Colab using Python libraries such as Pandas and BeautifulSoup, ensuring accurate and structured comparisons. By employing basic string matching techniques, the tool aligns similar products and displays prices in an easy-to-read tabular format, highlighting the best deals. To enhance user experience, the project incorporates data visualization techniques using Matplotlib and Seaborn, allowing users to quickly identify price variations and trends. The tool eliminates the need for complex API integrations, making it accessible to users with minimal setup. Additionally, it supports dynamic data updates, enabling users to modify or expand the dataset with new products and pricing information. This project serves as an educational demonstration of data processing, web scraping, and analysis using Python, offering a practical example of real-world decision-making in e-commerce.

## **1.Introduction**

In today's fast-paced digital marketplace, online shopping has become an integral part of consumers' lives. With the vast number of e-commerce platforms available, finding the best price for a product can be a challenging and time-consuming task. Customers often need to browse multiple websites, compare prices manually, and ensure they are getting the best possible deal. This project aims to address this challenge by developing a simple yet effective price comparison tool that enables users to compare product prices across multiple e-commerce platforms efficiently. The system leverages the power of Python and Google Colab to facilitate seamless data collection, processing, and visualization. It from a pre-collected or lightly scraped dataset. To ensure accuracy and usability, the project employs data processing techniques using Pandas and BeautifulSoup, aligning similar products based on basic string matching methodologies. The results are then presented in a structured tabular format, making it easier for users to identify the best deals available.

## 2.Realated work

The development of price comparison systems for e-commerce platforms has evolved significantly with the integration of machine learning and datadriven techniques. Early solutions primarily relied on rule-based scraping tools such as BeautifulSoup and Scrapy to gather product data from online retailers. However, due to the dynamic nature of web structures and anti-scraping mechanisms, these traditional approaches faced limitations in scalability and accuracy.

## 3.Methodolohy

3.1 System Architecture



#### 3.2 Modules

Module 1: Data collection Module 2: Data Pre processing Module 3: Model Implementation (Machine Learning - Random Forest) Module 4: Data Management Module 5: Graph Visualization

## 4. Experimental Results

#### 4.1 implementation Environment

Languages/Tools:python,webscraping,BeautifulSoup Hardware:Intel Core Duo 2.0 GHz or higher Software: Visual studio code + browser

#### 4.2 Test case & results

- 1. Add Product with Valid Data
- Ensure a new product with correct details is saved and appears in the product list. 2. Scrape Prices from All E-commerce Sites
- Confirm the scraper/API updates product prices correctly and stores them in the database. 3. Compare Prices Across Multiple Sites
- Verify that the system displays correct lowest, highest, and average prices per product.
  Train ML Model with Historical Price Data
- Test if the ML model trains without error and stores model version and metrics.
- Predict Future Price Using ML Ensure the system generates and stores a valid predicted price for a given product.
- Handle Site Not Reachable During Scraping
- Check that scraper handles unreachable or failed site connections gracefully (e.g., logs error).
- 7. Display Price Comparison on Product Page
- Confirm the UI shows price data from different sites clearly and accurately.
- Insert Invalid Foreign Key (Database Integrity Test) Ensure the database rejects invalid foreign key references (e.g., non-existent product\_id).
- Prevent SQL Injection on Search Field Verify input sanitization to prevent malicious SQL injection attempts.
- **10. Predict and Show ML Price with Timestamp** Check that every ML prediction is stored with a date and model version in the ML\_Predictions table.

### 5. Conclusion and future work

This project successfully demonstrates the implementation of a price comparison tool using Google Colab, Python, and machine learning techniques. By leveraging web scraping, data preprocessing, and visualization tools, the system enables users to compare product prices from multiple e-commerce platforms efficiently.

#### References

- 1. Chen, F. (2024). Research on Real-time E-commerce Price Comparison System Using Python Web Scraping Technology. International Journal of Computer Science and Information Technology, 4(2), 127-136. 🗆 cite 🗆 turn0search3 🗆
- Xiao, S., Wang, H., Ling, Z., Wang, L., & Tang, Z. (2020). Sentiment Analysis For Product Reviews Based on Deep Learning. Journal of Physics: Conference Series, 1651(1), 012103.
- 3. Anil, B., Singh, A. K., & Thota, S. (2018). Sentiment Analysis for Product Reviews. International Journal of Advanced Research in Computer Science, 9(Special Issue 3).
- 4. Sivaganesan, D., Aggarwal, D., Sridhar, K., & Arunkumar, M. (2022). Feature-based Sentiment Analysis for Product Reviews. International Journal of Engineering Research & Technology (IJERT), 11(06). □ cite□urn0search4□
- Yahav, I., Shehory, O., & Schwartz, D. (2018). Comments Mining with TF-IDF: The Inherent Bias and Its Removal. IEEE Transactions on Knowledge and Data Engineering, 31(3), 437–450.
- Sun, X., Ye, J., & Ren, F. (2014). Multi-strategy Based Sina Microblog Data Acquisition for Opinion Mining. International Conference on Intelligent Computing (ICIC), 551–560.
- Hu, N., Bose, I., Koh, N. S., & Liu, L. (2012). Manipulation of Online Reviews: An Analysis of Ratings, Readability, and Sentiments. Decision Support Systems, 52(3), 674–684.