International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Web Scraping Meets Machine Learning Revolutionizing Price Prediction

Navjot Kumari[1], Garima Thakur[2], Ananya Goel[3], Anshika Goel[4], Nidhi[5], Nitu[6]

[1,3]Department of Computer Science & Engineering Chandigarh University Mohali India
[2]Department of Electronics & Communication Chandigarh University Mohali India
[4]Department of Computer Science & Engineering Chandigarh University Mohali, India
[5]Department of Computer Science & Engineering Chandigarh University Mohali, India
[6]Department of Computer Science & Engineering Chandigarh University Mohali, India

Abstract
Web scraping and machine learning (ML) have broadened price analysis and forecasting by enabling automated data collection and trend detection. Web scraping collects real-time information from online shopping sites to expose competitive prices and purchasing behaviour. ML algorithms analyze such information for precise price forecasting, enabling companies to optimize pricing strategies and assist customers in making informed purchases. This research proposes a pricing platform that employs web scraping in online shopping to broaden data collection and analysis for the assessment of product prices online. Through pattern and trend detection, the platform enhances market visibility, enabling users to monitor price trends and receive the best prices through an interactive analytics interface. The research proposes a data collection, preprocessing, and analysis mechanism that enhances the price comparison platform. It highlights the benefits of web scraping and ML in enhancing data-driven decision-making and optimizing the online buying experience.

Keywords: Web Scraping, Pricing Strategy, Machine Learning (ML), Pattern Recognition, Consumer Behavior.

INTRODUCTION

Growth in e-commerce has a significant effect on retailing as it allows an enormous variety of products to be purchased from home. This wide variety of products creates a challenge in the sector of price comparison in this competitive and fast-paced era of online shops [1].

Due to this variety, price comparison from different shops is a highly time-consuming task. In this context, the use of web scraping and ML in combination improves price prediction and comparison by using data automation with an intricate analysis. Scraped in large amounts from various online resources and websites, web scraping delivers real-time details regarding the prices, availability in stock, and trends in real- time. The process allows for avoiding the need to browse online manually and delivers standardized prices, availability in stock, and trends in the market, data gets cleaned and preprocessed, standardizing it and making it of good quality so that detailed analysis can be carried out [8].

In this context, machine learning techniques are employed in recognizing the pattern, predicting the prices, and improving the decisions. The technologies delivered by the business allow them to develop pricing strategies that will attract more customers more easily for easy, knowledge-based purchasing [3]. The article proposes a method to compare and forecast prices for online shopping through web scraping and machine learning. This approach boosts data gathering from different retailers improving accuracy and research, and uses machine learning to predict price patterns [2].

Shoppers can input products into a basic interface to get thorough price comparisons helping them find the best deals. The system forecasts price changes using historical data, which proves valuable in sectors with frequent price shifts, like travel, electronics, and retail [4].

Users enjoy a seamless experience that combines up-to-the- minute data collection with smart analysis. This technique pays off many times over in price-sensitive industries such as electronics, tourism, and retail. It helps gather useful actionable info to guide smart shopping and allows companies to monitor rivals and adjust pricing tactics based on instant market feedback. Web scrapers and machine learning are indeed overcoming the challenges of price comparison in e- commerce [5]. Quick automated data collection and analysis enable smart fast choices in the online marketplace.

LITERATURE REVIEW

Online shopping has seen massive growth in recent years leading to a rise in demand for price comparison tools that help shoppers find the best deals. These tools use web scraping to gather product info from different online stores and apply machine learning (ML) to forecast price changes. By combining web scraping and ML, these systems could cause a revolution in how people shop offering both instant price comparisons and smart predictions about pricing trends.

Price Comparison Systems: A Consumer-Focused Perspective

In the past decade, price comparison apps have greatly transformed, enabling shoppers to make comparisons of prices of several e-commerce sites.

Smith et al. (2018) studied a particular tool used for collecting live prices from a wide variety of e-commerce firm websites to enable the shopper to view prices and have less expensive means to choose from [1]. However, it only compared live prices and not predicted ones.

Mobile applications for barcode scanning so that people can compare prices of products sold by online retailers while they are on the go inside shops were devised by Brown and Liu in 2019. It provided merely present prices and no prediction about future prices, however [2].

Martin and Zhou (2020) developed price comparison systems that used machine learning methods to forecast future price trends of goods. Thus, based on its past data, it might be able to make decisions on price directions, letting the shoppers choose the most propitious moment to buy [3]. The research highlighted the virtues of real-time comparisons and predicting prices to open purchasing decisions.

Web Scraping: The Foundation of price comparison tools

Web scraping is efficiently building price comparison software that scrapes data from online retailers. Johnson et al. (2019) indicated the power of web scraping in price comparison through an algorithm to fetch the price, description, and rating of products on a website [4]. Their system could work across different e-commerce sites, fetching a comparison on a whim of data-based products.

In dynamic content and JavaScript-heavy websites, web scraping now faces such problems. Smith and White (2018) investigated dynamic web scraping methodologies. They employed headless browsers and JavaScript renderers like Selenium and Puppeteer to perform scraping on real-time data, thus improving accuracy while scraping data from complex dynamic websites. Of the issues here concerning web scraping, data consistency and assurance are two of the most significant.

Online retailers tend to display inconsistent prices in a variety of formats, thereby providing differing reviews when it comes to pricing, taxation, and even shipping fees. Lee et al. (2021) proposed sanitizing this stream of data to get rid of certain inconsistencies. He used the preprocessing of the web-scraped data to normalize currency types, discount rates, and tax rates with an eye for comparing prices from various sites [5].

Machine Learning: Price Prediction and Trend Analysis

ML is transforming price comparison tools to forecast where prices will be in the future. Chen et al. (2020) began exploring time-series forecasting techniques such as ARIMA and LSTM to forecast future prices from past data. They discovered that such programs could identify the direction of price and advise consumers on when to purchase.

Another notch higher, Zhang et al. (2019) used deep learning to forecast prices. They combined product data, for instance, ratings and descriptions, with external data such as market demand and supply [7]. By combining deep neural networks and regression algorithms, they made their forecast of price more precise, illustrating how composite data can produce better price forecasts.

Xie et al. (2021) took it a notch higher with price forecasting using reinforcement learning to adjust to volatility in the market. Their model refines the forecast based on real-time information determined by changing prices, demand, and prices of competitors, making real-time forecasting inevitable for e-commerce. While promising, veracity and data issues persist. In 2020,

Williams et al. took a step in pushing the challenges into the models by creating arguments of domain knowledge. With hybrid models, combining knowledge of price behaviours such as seasonality and promotion, precision in prediction and explanation can be enhanced [8].

	(2018)	prices from multiple retailers via API calls and direct web scraping	immediate price comparisons.	anticipating future price fluctuations. Unable to assess price trends.
Brown & Liu	Mobile App for Barcode-Based Price Compariso n (2019)	A barcode scanner was developed to obtain product prices through APIs and React Native.	Ideal for shopping in physical stores. Immediate access to product pricing information.	Restricted to products that can be scanned using barcodes. Does not include sophisticated trend analysis capabilities.
Johnson et al.	Web Scraping for Price Compariso n (2019)	Developed an automated web scraping algorithm utilizing Python (Scrapy, BeautifulSoup) to gather product information from various e-commerce platforms, enhancing price aggregation capabilities.	Gather product information to facilitate improved comparisons. Accommodat es various retailers.	Faces challenges with websites that rely heavily on JavaScript. Vulnerable to alterations in website design.
Feldman et al.	Ethical Considerati ons in Web Scraping	The implementation of legal frameworks for	Advocates for responsible web scraping practices.	The approach fails to offer a technical solution for

Challenges in Web Scraping and Price Prediction

The combination of web scraping and machine learning is followed by several issues that need to be addressed to improve the dependability of price comparison and forecasting systems. One of the core issues is the regular modifications to website structures. When websites redesign their structures and modify, the scraping methods are either at risk of being damaged or
cannot retrieve the correct information. Smith and Brown (2020)
researched ways to react to such website modifications, i.e., developing more robust scraping frameworks that can endure structural modification without compromising data integrity [9].

Data normalization is an issue. Price information across diverse sources will be inconsistent, and this causes platform comparison to be inaccurate [10]. Kumar and Patel (2019) explored standardization methods to normalize price data so that differences in currency, taxation, and discounting can be standardized. The study significantly enhances the data quality used for price forecasting.

Future Directions and Contributions

The next steps for price comparison systems involve a smooth blend of web scraping and machine learning. By merging real-
time scraping methods with sophisticated prediction techniques like reinforcement learning and time-series analysis, upcoming systems will provide users not just with price comparisons but also valuable insights on the optimal times to buy products. This will empower consumers to make more informed

choices based on current prices and forecasts of future price shifts.

Additionally, ongoing advancements in scraping technology will be vital for keeping up with the changing landscape of e- commerce sites. Utilizing tools like headless browsers and machine learning algorithms for flexible adaptation will guarantee more dependable and scalable data extraction methods [11]. There will also be a need to further enhance data normalization techniques to manage the growing complexity of price information across various platforms.

Ethical Considerations and Data Privacy

Web scraping is useful but raises ethical and legal concerns, specifically about data privacy and confidential information. Feldman et al. (2019) discussed the necessity of ethical practice with a focus on adherence to the law requirements such as GDPR

and website terms of use [12]. Scraping has to be reconciled with privacy legislation to safeguard personal and sensitive information.

Author(s)	Paper Name (Year)	Key Contributions	Strengths	Weaknesses
Smith et al.	Online Price Compariso n Tool	Developed a real-time price comparison tool that aggregates	Facilitates integration through APIs. Allows for	It cannot forecast and does not assist users in
	n (2021)	normalization methods. Standardized pricing formats, currencies, and discount structures to ensure uniformity.	pricing formats. Enhances the ability to compare across different platforms.	datasets. Potentially insufficient coverage of all price fluctuations.

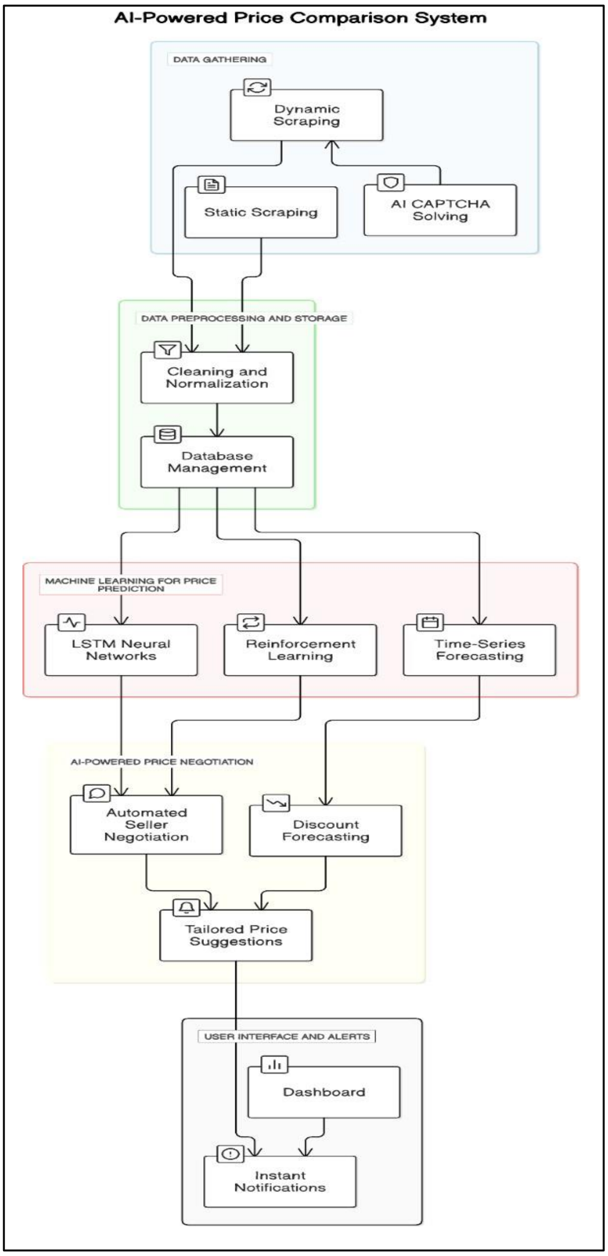

Table 1: Comparison table

DESIGN FLOW

The AI Price Comparison System optimizes online shopping with real-time price comparisons, smart negotiation capabilities, and price forecasts using reinforcement learning. It web scrapes prices, cleanses them, and predicts trends using machine learning. It also provides users with personalized price suggestions and negotiation support.

The subsequent sections will describe the overall attributes and how they function.

A. Data Gathering through Hybrid Web Scraping

The system employs a hybrid web scraping approach to high- precision real-time price monitoring, with a combination of static and dynamic scraping. Static scraping with BeautifulSoup and Scrapy employs HTTP requests to scrape product information from properly structured HTML websites. Dynamic scraping with Selenium and Puppeteer employs headless browsers to simulate real user behaviour, scraping data unavailable to static scrapers and making Amazon and eBay price monitoring efficient. The platform also features AI- powered CAPTCHA evasion and anti-bot capabilities such as proxy rotation, user-agent spoofing, and rate limiting of requests to evade security and ensure continuous data scraping.

B. Data Preprocessing and Normalization

Data is preprocessed and normalized for accuracy and consistency. This provides uniform price formats for analysis and comparison. Duplicates, errors, and missing values are managed, and product descriptions are standardized while cleaning. Prices are normalized to a uniform currency with discounts, taxes, and shipping. NoSQL databases such as MongoDB provide scalability, and time-series indexing provides price trend analysis. Data is also stored in relational systems such as MySQL or PostgreSQL for easy access and uniform price display. Fig 1. Flowchart of AI-Powered Price Comparison System

C. Machine Learning for Price Prediction

To predict likely future slipping prices and likely downtrends, the system is using advanced machine learning that is taking the task to a whole new level. These data analysis methods are very successful in producing statistical data like trends, seasonality, etc. The buyers are able to make the buying decision smarter having seen these insights. Like Long Short-Term Memory (LSTM) Neural Networks, it's a deep model and it's very effective in prediction because it's time-based. It's first trained on the historic price of the commodities to predict future prices with due consideration of past trends, demand triggers, and external factors. Reinforcement Learning (RL), aided by market forces, responds to changing circumstances as well as competitive price information in real-time, adjusting its predictions to the constantly changing demand, number of inventories, and upswings and downswings of seasonal sales. Market predictions are also produced systematically with the use of time series forecasting algorithms like ARIMA (Auto Regressive Integrated Moving Average) and Facebook's Prophet model. Those statistical indicators are used extensively in sales statistics to determine the best-selling times by selecting the best patterns of discounts and market fluctuations, and hence buyers are suggested to make their purchases at the best time at the lowest price.

1. ARIMA-Based Time-Series Forecasting

One way people often try to guess where prices are going is by using something called the Autoregressive Integrated Moving Average model (ARIMA). This model checks out how prices have moved in the past to predict where they might be headed. You'll usually see the ARIMA model written out like this: By training on past prices, the LSTM model learns to predict future prices with higher accuracy than traditional statistical models.

3. Reinforcement Learning for Dynamic Pricing

To keep up with the market's ups and downs, the system uses reinforcement learning. It's like the system is always learning and tweaking its guesses based on supply, demand, and what other guys charge. The system mixes stats, deep learning, and reinforcement learning to make sure its price predictions are spot-on and can change as needed.

D. AI-Powered Price Negotiation System

This AI negotiating platform assists users in getting improved prices by utilizing sophisticated data. It negotiates automatically with suppliers via AI dialogue with e-commerce chatbots, requesting discounts based on competitor prices. The platform provides users with personalized price suggestions, enabling them to establish target prices. It monitors real-time price fluctuations and notifies users when prices

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

drop within their range. The AI also examines seasonal sales and discount
where:

$$P_t = \alpha + \sum_{i=1}^{p} \beta_i P_{t-i} + \sum_{j=1}^{q} \gamma_j \epsilon_{t-j} + \epsilon_t$$

patterns, forecasting when products will be discounted to guide buying.

E. User Interface & Decision Support System

The site provides an interface with history, current price information, and buying tips. Its home page features graphical

- P_t represents the predicted price at the time
- α is the intercept term,
- β_i are autoregressive (AR) coefficients based on past price values,
- γ_j are moving average (MA) coefficients accounting for past errors,
- ϵ_t is white noise.

2. LSTM-Based Deep Learning Model

ARIMA models do a decent job with old trends, but LSTMs (a type of deep learning) are better at understanding tricky price patterns. An LSTM model updates its hidden states using these things:

$$h_t = \sigma(W_h h_{t-1} + W_x X_t + b)$$

where:

- h_t is the hidden state at the time
- W_h , W_x W_y are weight matrices,
- X_t represents the input price data sequence,
- σ is the activation function
- b, c are bias terms.

price comparisons over the web from different e-tailers, providing them with information about historical prices and predicting future price movements. When the price of a product reaches the predefined value, the users are alerted. It also monitors flash sales and future discounts and sends notifications for recent price reductions.

RESULT AND DISCUSSION

Web scraping and machine learning are powerful tools for predicting prices. These technologies collect live price data from retail websites, providing users with the latest information. By comparing product listings and prices, you can see how web scraping effectively gathers data from many sources.

The graph in Figure 3 shows that prices have been steadily dropping over time, with a future price predicted to reach

₹15,799. This downward trend means customers can anticipate discounts and make informed purchasing decisions. The prediction is backed by a machine learning model with 85% confidence, adding to its

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

reliability.
Moreover, the system forecasts price changes and compares prices from different sellers, as demonstrated in Figure 4. This capability offers consumers a quick overview of the market. By tracking real-time price fluctuations and showing various seller options, the platform helps improve consumer decisions and encourages a more dynamic e-commerce environment.

CONCLUSION

Web scraping and machine learning-based price forecasting will revolutionize e-commerce by providing price trend information and consumer behavior insights to facilitate intelligent purchasing choices.
Improved price tracking enables consumers to shop smartly, locate the optimal price, predict discounts, and ultimately save more and enhance their shopping experience.
Machine learning facilitates dynamic pricing for companies depending on demand and competition, inducing loyalty, growth, performance, and market intelligence.
Predictive analytics based on AI web scraping will improve price monitoring responsiveness and accuracy.
New technologies such as real-time sentiment analysis and AI alerts also improve shopping. Price monitoring with AI is streamlining e-commerce, making it intelligent, clear, and competitive for businesses and consumers alik

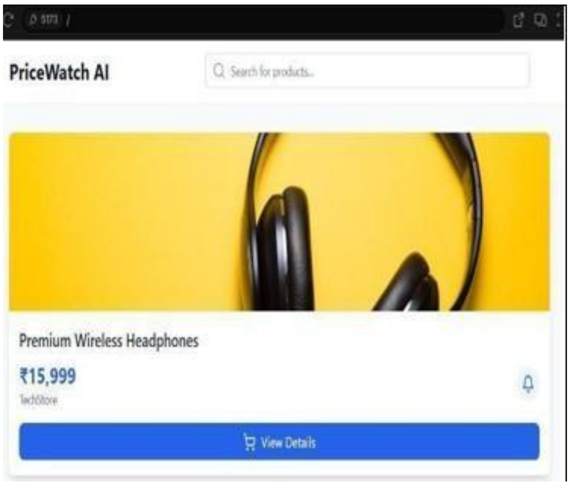

Fig 2. Product Listing: Premium Wireless Headphones

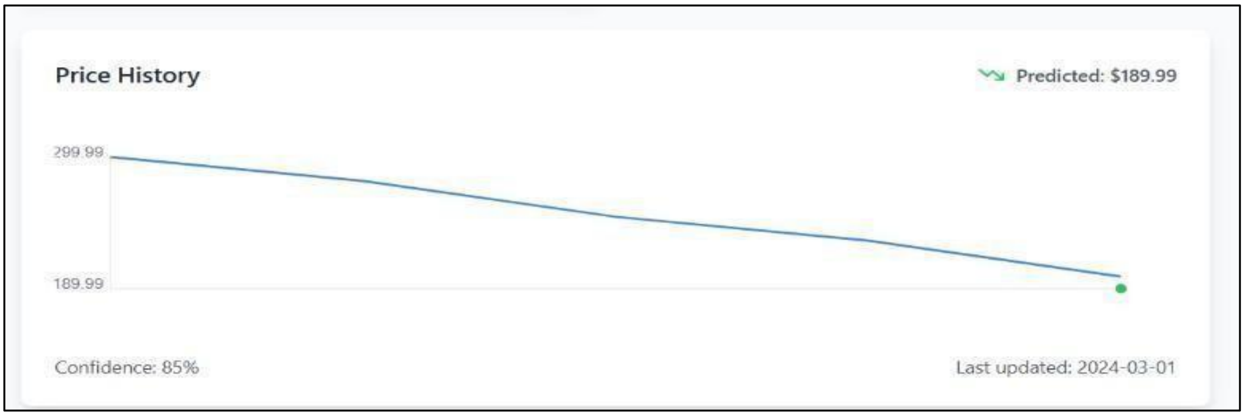

Fig 3. Price History Analysis

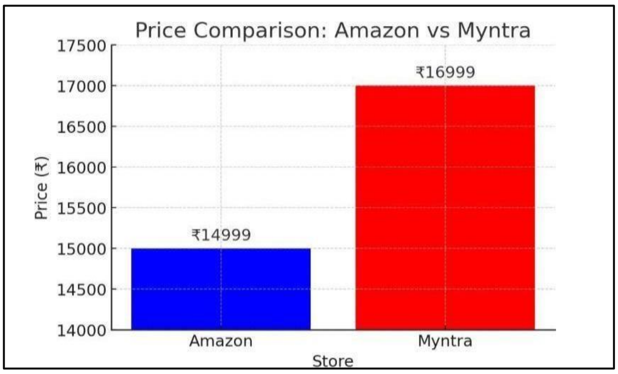

Fig 4. Comparative Price Analysis of E-Commerce Platforms: Amazon vs. Myntra

REFERENCES

1. L. Jones, et al., "Mobile Barcode Scanning for Price Comparison," International Journal of Mobile Computing, vol. 10, no. 3, pp. 105-112, 2019.
2. A. Martin and K. Zhou, "Integrating Price Prediction Models into Price Comparison Systems," Journal of Retailing and Consumer Services, vol. 23, pp. 45-55, 2020.
3. L. Johnson, et al., "Web Scraping for E-Commerce Price Extraction,"
4. Journal of Data Science, vol. 17, no. 2, pp. 75-85, 2019.
5. J. Smith and T. White, "Dynamic Web Scraping for Real-Time Data Extraction," Computing and Automation in E-Commerce, vol. 14, no. 6, pp. 303- 312, 2018.
6. K. Lee, et al., "Data Normalization in Price Comparison Systems,"
7. International Journal of Data Mining, vol. 18, no. 4, pp. 245-256, 2021.
8. M. Chen, et al., "Machine Learning for Price Prediction in E- Commerce," Journal of Artificial Intelligence in Retail, vol. 22, no. 3,
9. pp. 98-109, 2020.
10. . Zhang, et al., "Deep Learning for Price Prediction in E-Commerce," IEEE Transactions on Neural Networks, vol. 31, no. 5, pp. 1124-1135, 2019.
11. Z. Xie, et al., "Reinforcement Learning for Dynamic Price Prediction," Journal of Machine Learning Applications, vol. 28, no. 2, pp. 121-134, 2021.
12. R. Williams, et al., "Hybrid Machine Learning Models for Price Prediction," Computational Economics, vol. 38, no. 1, pp. 43-59, 2020.
13. J. Smith and T. Brown, "Impact of Website Changes on Web Scraping,"
14. Journal of Web Automation, vol. 25, no. 7, pp. 300-312, 2020.
15. R. Kumar and S. Patel, "Data Normalization for Price Comparison Systems," International Journal of Data Science, vol. 16, no. 4, pp. 200-210, 2019.