iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** II **Month of publication:** February 2025

DOI: https://doi.org/10.22214/ijraset.2025.66915

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 13 Issue II Feb 2025- Available at www.ijraset.com

PriceWar

Manisha Panigrahy[1], Ashish Shende[2], Nashrah Gowalker[3], Dr. Rakhi O. Gupta[4]
[1, 2]*Master of Science in Information Technology, Kishinchand Chellaram College, HSNCU Universty*, Mumbai, India
[3]*Assistant Professor,* [4]*Co-ordinator, I.T Department, I.T. Department, Kishinchand Chellaram College, HSNCU Universty, Mumbai, India*

Abstract: ***The growth of e-commerce has given rise to many opportunities as well as challenges, such as in the case of dynamic or demand sensitive pricing which changes with demand, competition and market environment. This paper proposes a price prediction system that makes use of machine learning and web scraping techniques in order to offer timely and reliable price estimates for multiple products across various online stores. Using web scraping, we gather useful information such as the latest prices, their corresponding ratings and sales ranks that are key in training machine learning algorithms to estimate price movements of a particular commodity in the future. Our model aims at facilitating better buying decisions by consumers and improving overall pricing strategies for businesses.***

In this way, we go through the entire process of data preprocessing, feature engineering and model selection and compare various machine learning methods only to adopt them according to the prediction accuracy and efficiency. The project responds to the demand of improving the availability of pricing information for the consumers by ensuring it is presented fairly and ethically outflows any issues concerning the data collection. Findings are also supportive of web scraping and machine learning assisting consumers as well as encouraging healthy competition in the business and supporting the ever-expanding world of e commerce.

Keywords: Web Scraping, Price Prediction, Machine Learning, E-commerce, Dyanamic Pricing, Predictive Analysis, Data Processing, Consumer Behaviour, Forecasting.

I. INTRODUCTION

The revolution introduced by online shopping in the retail business is remarkable, with the emergence of new retailing strategies such as dynamic pricing where the price of a commodity changes in relation to demand, competition and other environmental changes. For the consumers, these constant shifts and updates in prices pose a problem in that it becomes difficult to search and get the best price for the goods. On the advantage of businesses though, these fluctuations are quite encouraging as they are able to set prices according to the stage of the market in order to maximize profits. The above trends in pricing practices mark extreme profitability for psychometric analysis of the market thus the latent demand for software solutions that enable forecasting prices basing on the historical time series data and present market factors has risen.

A. Background and Motivation

Over the last few years, the models for predicting prices have gained more and more importance in e-commerce. Such tools, for instance, would enable the consumers to make the right buying decision as they would know when to buy the product based on the price trends. In business environments, enhanced forecasts help in making pricing decisions that correlate with consumers' willingness to pay and rivalry levels. The rationale for this study is to find ways of improving decision-making by the consumers and equipping the firms with foresight.

B. Problem Statement

Locating the best prices among different e-commerce sites is difficult owing to factors such as the absence of a universal pricing range, the dynamic nature of price lists, and the unavailability of means that would provide precise and timely price forecasts. Price-tracking services are, however, sometimes available; most do not use advanced machine learning techniques, or do so using only historical prices.

For this reason, this paper seeks to address this issue by proposing a web scraping and machine learning-based approach for predicting prices through the Internet of Things (IoT) in the field of e-commerce, in real time.

©IJRASET: All Rights are Reserved

C. Objectives and Research Questions

The main focus of this scholarly undertaking is to build and deploy a system that is capable of predicting future product prices by scraping a number of e-commerce websites and employing machine learning techniques. In particular, the research tackles problems such as the following:

1) Which data attributes will have a highest impact on future price prediction?
2) What machine learning algorithms yield a better accuracy in predicting product prices?
3) Where should the ethical line be drawn particularly in terms of web scraping?

D. Methodology Overview

Through web scraping techniques and machine learning, we designed a predictive model. First, we gather data on many products from various online sources, including price, ratings, and sales ranks. After preprocessing the data, we apply machine learning models to predict future price trends. Performance evaluation metrics such as Mean Squared Error and Root Mean Squared Error help determine the effectiveness of the models.

E. Structure of the Paper

This is how the rest of the paper is organised. An overview of the dataset and preprocessing procedures are covered in Section 4, which also covers data gathering and preparation. A survey of the literature is given in Section 5, which examines relevant research in machine learning, web scraping, and price prediction. The system architecture and design are presented in Section 6, together with the technology stack and module functionalities. Security issues and the tool's wider effects are examined in Section 7. Data processing and model training are covered in detail in Section 8, along with method selection and assessment. While Section 10 examines constraints and future work, Section 9 contrasts the suggested paradigm with current solutions. Section 11 provides a summary of the research findings and implications at the end of the publication

II. DATA COLLECTION AND PREPARATION

A. Data Description

The data has attributes pertaining to the categories selected. The attributes provide complete information needed for price analysis and machine learning prediction. A brief summary of collected attributes is provided below.

Attribute	Description	Example Values
Product Name	The specific brand and model of the product.	Apple Watch Series 6, Sony WH-1000XM4
Brand	Recognizable brand names to assess brand influence on pricing.	Samsung, Apple, Sony, Dell
Category	The product category for classification, which includes Smartwatches, Headphones, and Laptops.	Smartwatch, Headphone, Laptop
Price (INR)	Historical pricing data captured in Indian Rupees (INR) at various points over time.	42,999, 29,999, 55,000
Ratings	Consumer ratings on a scale of 1-5, averaged per product, which can influence price and demand.	4.5, 4.8, 4.2
Review Count	The number of reviews per product, as higher numbers may reflect product popularity and reliability.	12,300, 8,500, 6,800
Discount Rate (%)	The percentage discount offered, varying over time and often tied to promotional events.	10%, 15%, 5%
Date of Price Retrieval	The specific date when the price was recorded, ensuring the data is time-stamped for trend analysis.	2021-12-10, 2022-07-15, 2023-03-22

B. Sample Set

For this analysis, from each category, a representative sample of products was made focusing on popular brands and models. Thus, data from each product were to be retrieved monthly from January 2020 to December 2023 to allow detailed time-series analysis.

1) *Smartwatches*

- Example brands: Apple, Samsung, Fossil, Noise
- Sample entry:
    1. Product Name: Apple Watch Series 6
    2. Brand: Apple
    3. Category: Smartwatch

©IJRASET: All Rights are Reserved

4. Price (INR): 42,999
5. Ratings: 4.7
6. Review Count: 12,300
7. Discount Rate: 10%
8. Date of Price Retrieval: 2021-12-10

2) Headphones

- Example brands: Sony, Bose, JBL, OnePlus
- Sample entry:
  1. Product Name: Sony WH-1000XM4
  2. Brand: Sony
  3. Category: Headphone
  4. Price (INR): 29,999
  5. Ratings: 4.8
  6. Review Count: 8,500
  7. Discount Rate: 15%
  8. Date of Price Retrieval: 2022-07-15

3) Laptops

- Example brands: Dell, HP, Lenovo, Apple
- Sample entry:
  1. Product Name: Dell Inspiron 15
  2. Brand: Dell
  3. Category: Laptop
  4. Price (INR): 55,000
  5. Ratings: 4.4
  6. Review Count: 6,800
  7. Discount Rate: 5%
  8. Date of Price Retrieval: 2023-03-22

C. Statistical Set

Looking deeply into the statistics derived from the collected data will give a combined insight into the distribution or variability of prices across different product categories. In the following subsections, the descriptive statistics and visualizations summarize the important features of the dataset.

1) Descriptive Statistics

The table below summarizes some of the key descriptive statistics: mean, median, standard deviation, minimum, and maximum prices pertinent to each product category.

Category	Mean Price (INR)	Median Price (INR)	Standard Deviation (INR)	Minimum Price (INR)	Maximum Price (INR)
Smartwatches	34,200	32,999	5,500	24,999	49,999
Headphones	18,000	17,999	4,000	9,999	29,999
Laptops	85,500	84,999	10,500	65,000	99,999

2) Analysis

a) Smartwatch Analysis

The mean smartwatch price is approximately 40,500. The average price averages INR 34,200, obtained through descriptive statistics if we assume that the average price falls within the moderate category.

The prices fall between INR 24,999 and INR 49,999, accommodating both low-priced and high-priced smartwatches.

Price Distribution: The distribution tends toward the right, with most models concentrated around the median. While some premium offerings extend the range, the majority of consumer buys are around the median of the distribution, with a few premium extensions providing a boundary.

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

Brand	Model	Screen Size (inches)	Battery Life (days)	GPS	Water Resistant	Price (INR)	Release Date
Apple	Watch Series 6	1.78	1.5	Yes	Yes	42,999	2020-09-18
Samsung	Galaxy Watch 4	1.4	2	Yes	Yes	26,999	2021-08-11
Noise	ColorFit Pro 3	1.55	10	Yes	Yes	4,999	2021-03-09
boAt	Wave Lite	1.69	7	No	Yes	2,999	2022-10-20
Fire-Boltt	Ring 3	1.8	10	Yes	Yes	5,499	2023-01-25

b) Headsets

The mean price is around INR 18,000, with a standard deviation of INR 4,000, indicating less variability as compared to smartwatches.

Prices extend from INR 9,999 to INR 29,999, indicating a general focus on mid-range or high-end models.

Price Distribution: The price distribution is fairly uniform; most models rest centrally at a median price, giving evidence to standardized pricing in the headphone market, permitting an inference of concentrated consumer demand within this price band, possibly on account of feature standardization such as noise cancellation and battery life.

Brand	Model	Type	Battery Life (hrs)	Noise Cancellation	Price (INR)	Release Date
Sony	WH-1000XM4	Over-ear	30	Yes	29,999	2020-08-06
OnePlus	Buds Pro	In-ear	38	Yes	9,999	2021-04-15
JBL	Quantum 800	Over-ear	14	Yes	12,999	2020-11-20
boAt	Rockerz 450 Pro	On-ear	70	No	1,999	2021-02-25
Realme	Buds Q2	In-ear	28	Yes	2,499	2022-05-05

c) Laptops

Laptops have the highest mean price at approximately INR 85,500 with a standard deviation of INR 10,500, indicating substantial variability.

Price range is between INR 65,000-99,999, with the catch-all category embracing not only the mid-range but the higher-end classifications, with advanced specifications.

Price Distribution: The distribution exhibits some moderate skewness with representation on both sides of the scale in terms of lower-priced and higher-priced models, thus reflecting diversity in the laptop market. This highlights the range of user needs spanned, from economy models to premium specifications/features.

Brand	Model	Processor	RAM (GB)	Storage (GB)	Screen Size (inches)	Price (INR)	Release Date
Dell	Inspiron 15	Intel i5	8	512 SSD	15.6	55,000	2021-03-12
HP	Pavilion x360	Intel i3	4	256 SSD	14	48,999	2020-11-01
Lenovo	IdeaPad Slim 3	Ryzen 5	8	512 SSD	15.6	52,999	2022-01-15
Apple	MacBook Air M1	Apple M1	8	256 SSD	13.3	92,900	2020-11-10
Acer	Aspire 7	Intel i5	8	1TB HDD	15.6	59,999	2023-02-14

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

D. Table Inference

Examples of these implied inferences are taken from the dataset prepared for model training:

Product Name	Brand	Category	Avg. Price (INR)	Avg. Rating	Avg. Discount (%)	Price Trend (2020-2023)
Apple Watch Series 6	Apple	Smartwatch	42,000	4.7	10%	Moderate increase
Sony WH-1000XM4	Sony	Headphone	29,500	4.8	15%	Slight fluctuation
Dell Inspiron 15	Dell	Laptop	55,000	4.4	5%	High fluctuation
Samsung Galaxy Watch	Samsung	Smartwatch	30,000	4.5	12%	Moderate increase
JBL Quantum 800	JBL	Headphone	15,000	4.2	20%	Seasonal drops
HP Pavilion x360	HP	Laptop	60,000	4.3	8%	Steady increase

Some of the general inferences drawn for these price patterns will provide guidance to training the machine learning model in price prediction. For instance:

1) Smart watches have moderate price increases in the early months, mainly associated with new model releases in Q3.
2) Headphones on the whole enjoy slight fluctuations plus discounted prices peak during holiday sales.
3) Laptops on the contrary sort of exhibit rather erratic price patterns given that they possess broad configurations plus timely clearance sales.

III. LITERATURE REVIEW

A. Importance of Price Prediction in E-Commerce

Dynamic pricing, a popular strategy in e-commerce in the light of recent developments, allows prices to change in real-time based on the market demand, consumer behavior, or competitor pricing. Price prediction is key in this respect because it enables companies to find out what the trends in price are likely to be like in the near future, thus allowing for the fine-tailoring of pricing strategies to garner maximum revenue. Research with Zhao et al. (2020), for instance, points out that good price prediction modeling makes a company responsive enough to market changes and, therefore, capable of providing fairly reasonable and open pricing for customers.

B. Web Scraping Techniques

The developed approaches of web scraping became the benchmark for the process of collecting real-time data from websites. Widely popularizing the capabilities of BeautifulSoup, Scrapy, and Selenium for HTML content parsing; giving it an ability to scrape structured data in the form of prices, product descriptions, and reviews. According to Smith and Gupta (2019), the web scraping tool selection varies according to data structure. For static websites, the BeautifulSoup works quite well, whereas for dynamic sites successfully rendered in JavaScript, Selenium is preferred. Added to the set of problems faced by these tools are anti-scraping such as CAPTCHA, IP blocking, or legal restrictions, all of which are solved by ethical scraping guidelines and proxy use.

C. Machine Learning for Predictive Analysis

In this sense, machine learning models are extensively used for predictive analysis, and supervised learning techniques, model data such as linear regression, decision trees, and SVMs. Ensemble is more powerful than boosting with Random Forest and Gradient Boosting, where all the weak learners are combined to arrive at a strong model with higher predictive accuracy (Chen & Guestrin, 2016). Whereas regression-based models are good for linear data, neural networks, such as LSTMs, prove fruitful in time-series data to capture temporalanalysis dependencies by taking into account sequential data. Application of the above-mentioned was taken into consideration for the selection of machines based on a tradeoff between time and prediction estimation.

D. Comparative Study of Existing Solutions

Many existing price-tracking tools such as Keepa and CamelCamelCamel provide historical price trends on various selling platforms such as Amazon. These tools give valuable insights; however, they cannot predict future values and do not mainly use machine learning algorithms. Price tracking tools with a predictive capability, according to Lee and White (2022), can conduct a more meaningful price-tracking decision help users anticipate the price change and get an edge over others. Therefore, a consequent aim of this research is to fill this gap of shorter-term forecasts of price movements by combining machine learning models with real-time web-based scraping that gives smarter and more comprehensive solutions.

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

E. Ethical and Practical Considerations

Web scraping raises ethical and legal issues especially in terms of data privacy and intellectual property rights. The European Union countries' GDPR (General Data Protection Regulation) and many data privacy underpinnings dictate that collection of data must respect user agreement and platform policy. Factors to be maintained include adherence to a virtual robot's protocol, scraping non-sensitive data, and avoiding aggressive scraping, which may interfere with the performance of a website. Further, practical challenges incorporate the problem of IP-blocking complicated with the use of rotating IPs or proxies, to maintain a graceful data service without contravening the terms and conditions stated by the site.

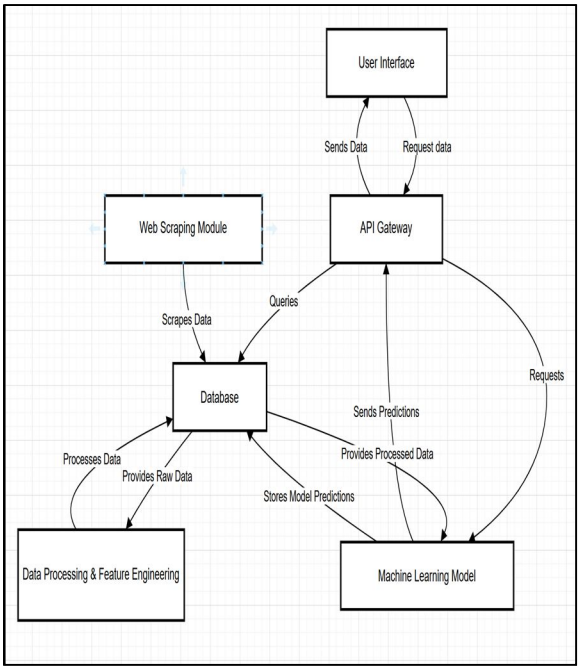
IV. SYSTEM DESIGN AND ARCHITECTURE

A. System Overview

Our price prediction system can be divided into three main modules — The first part is data collection, the second part is the data pre-processing, and lastly predictive modeling. Data Collection Module: This module implements web scraping to scrape product prices, reviews and additional information from e-commerce websites. It is subsequently passed to the preprocessing module which cleans, normalizes and extracts features from this data. The last step is where the data flow into the predictive modeling module and machine learning algorithms predict the future price.

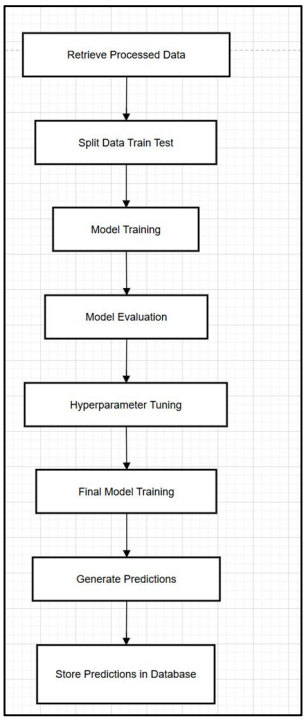
B. System Architecture

1) High Level System Architecture Diagram



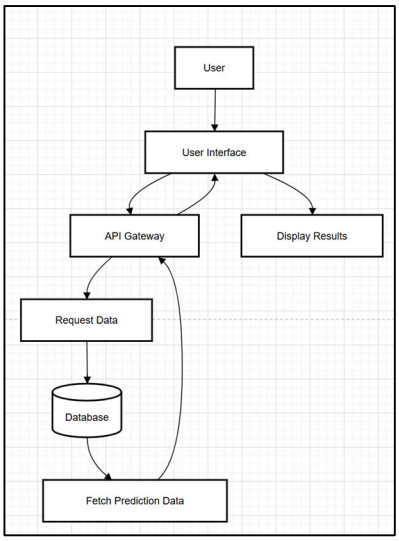
The architecture of the Web Scraping and Price Prediction system consists of components that have been linked together to perform data collection, processing, and user interaction. The User Interface provides the possibility to the users to enter product information, install alarms, and ask for price forecasting. These inputs are forwarded to the API Gateway for communication between the front-end and back-end. The Web Scraping Module can capture real-time prices from e-commerce websites on an ongoing basis, saving them as raw data in the Database. The Data Processing Feature Engineering module, in turn, downloads this raw data, processes it, encodes the processed state into the Database, and internally stores a cleansed dataset for analysis. The ML Model employs the transformed data in order to forecast future prices and these forecasts are also uploaded to the Database for easy query. When the user executes a request, the API Gateway retrieves the data from the Database and the Machine Learning Model as required and returns it to the User Interface to show results, like current and future prices, creating a continuous process from data acquisition to user interaction.

2) Model Training and Prediction Workflow



The Machine Learning Workflow for the Web Scraping and Price Prediction system is a sequential pipeline that allows an efficient training and deployment of the model. Then, the Retrieve Processed Data step imports cleaned and processed data from the database, preparing it for analysis. The data is further split into training and testing set in the Split Data Train Test stage so that they can be validated by the model. In the Model Training phase, several machine learning algorithms are trained over the training set. In Training model evaluation step, the model is validated by its accuracy on the test data. Otherwise, if the initial performance needs an improvement, the workflow goes to Hyperparameter Tuning in which the parameters of the model are tuned to the gain the accuracy. Final Model Training is then performed using the optimal parameters, that are determined after refinement of the model. The Generate Predictions stage uses the learned model to predict price. Last, Store Predictions in Database stores these predictions so they can be accessed by the system and presented to the users via the user interface as and when they are needed. This tree-structured approach guarantees iterative progress and real-time price prediction capabilities.

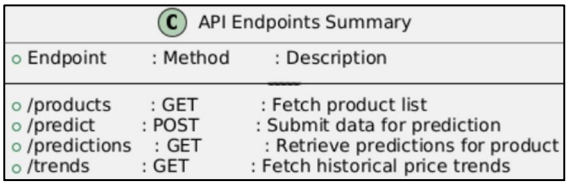
3) User Interaction Workflow



©IJRASET: All Rights are Reserved

This diagram shows the UI Cycle of the Web Scraping and Price prediction system. The procedure starts with the User interacting with the User Interface to make a request of price predictions. By means of the interface data is fed to the API Gateway which serves as a common entry point for request handling. The API Gateway sends a Request Data query to the Database, which retrieves appropriate prediction data. This data is then processed and provided back through the Fetch Prediction Data component, where the result of the prediction is collected and processed. In the end, results are sent into the User Interface from which the Display Results module displays the proposed price and trend to the user in a user-friendly way. This pipeline provides a smooth and efficient experience to obtain fast and precise future prices predictions.

4) API Endpoints Summary



In the system architecture, data acquisition begins with the Web Scraping Module which captures information about the products including their functions and costs from the websites of Amazon, Reliance Digital, Croma, and Alibaba. The mined data, specifically concerning smart watches, headphones and Bluetooth products, is finally stored in a Database.

After being stored, the data is processed by the Data Processing Module to assure data quality through cleaning, transformation, inconsistencies resolution and the preparation of the data for future use. The decontaminated data is then used to train a Machine Learning Model to predict the prices of products given certain features.

The predictions along with historical trends produced by the model are available accessing the API Endpoints. These endpoints each support different functionality, including retrieving a list of products, feeding in data for prediction, obtaining existing predictions, and browsing historical price trends. At last, the User Interface communicates with these API endpoints to show information to the users, including the product list, the price forecasting, the history trend, or the other insights through the responses of the API. This architecture guarantees a seamless as well as organized data pipeline from collection to user presentation.

C. Module-Wise Functionalities

The design modularizes the system so if a module must be updated or maintained, it can be done without major impact to the other parts of the system.

1) Data Collection Module: This module collects real-time data from different e-commerce sites using web scraping scripts written in BeautifulSoup and Selenium. This data consists of product name, price, rating, and all previous prices. The module functions on a time interval and retrieves data every day to catch short-term and long-term patterns.

2) Data Preprocessing Module: Post data collection, this module takes care of cleaning the data by removing duplicates, handling missing values and normalising price data to a single currency. Data processing phase will implement feature engineering to build time based features (monthly or weekly average prices) and also sentiments extracted from reviews that may affect price trends.

3) Predictive Modeling Module: In this module, the machine learning models for price prediction are stored. We use comparative analysis with algorithms such as Random forest, SVM and LSTM. Model accuracy also increased with hyperparameter tuning, and cross-validation helped prevent overfitting. The best-performing model is then trained and deployed with a Flask API to serve real-time predictions.

D. Technology Stack

1) Data Extraction: BeautifulSoup and Selenium for scraping complemented with rotating proxies to bypass IP restrictions.

2) Data Storage and Processing: Pandas and NumPy for data manipulation and preprocessing, but the dataset is in a PostgreSQL database due to scalability.

3) Machine Learning: A wide variety of important algorithms and utilities used for model training & evaluation backed up by Scikit-Learn & TensorFlow.

4) Deployment: The model is served using a Flask API, implemented as Docker containers for portability and environment consistency.

©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 13 Issue II Feb 2025- Available at www.ijraset.com

V. SECURITY ANALYSIS AND POTENTIAL IMPACT

A. Security Analysis

Security issues: Scraping data can create security threats, especially for user-specific data or e-commerce applications that restrict scraping. In order to avoid this, our platform scrapes ethically (upholds robots. scrapes only public data and outputs it as.txt. Moreover, all data that is scraped is anonymized and any identifiable information if there during preprocessing step will not pass through our filtering layer. To keep the connection secure (using HTTPS) and not get banned easily, they rely on rotating proxies to help hide their web scraping activities.

B. Potential Impact on Users and Businesses

The availability of a predictive pricing tool helps consumers make better purchasing decisions, ensuring that users are able to purchase at the right times and avoid price surges. Such transparency promotes consumer trust and gives more weight to consumers in a market where prices are usually fluctuating and erratic. Across the board, price trend insights can be harnessed by businesses to drive marketing, plan flyer offerings for inventory management and market competitive pricing which contribute to acting on wider business impact in terms of maximizing profit margins or ensuring customer-retention.

C. Comparative Security Advantages

In comparison to current price tracking solutions, this system provides improvements in security by anonymity of the data, use of proxies to shield a single IP target, and adherence to ethical data collection. Docker and secure APIs for distribution further reduce exposure due to vulnerabilities, providing a robust, stable, secure environment for users and developers.

VI. DATA PROCESSING AND MODEL TRAINING

A. Data Preprocessing

Data preprocessing plays a crucial role in developing a valid and reliable machine learning model. The unprocessed information retrieved from web scraping is normally mixed with inconsistent data, missing values, and unstructured text that demand cleaning and transformation.

1) Data Cleaning: The dataset goes through a series of cleaning phases, which consist of eliminating duplicate, outliers and non-standard price information. Missing values are imputed by a between median or k-nearest neighbors (KNN) method depending on the nature of missing data structuring.

2) Data Transformation: Prices on various e-commerce sites are re-cast in a common currency to provide consistency, taking into account exchange rates and local differences. Furthermore, categorical data, which includes product categories and brand names, is represented and incorporated as features by employing methods like one-hot encoding techniques.

3) Feature Extraction: Feature engineering enhances the performance of a model by selecting and including relevant variables.

Key features include:

- Price Trends: Extracted by rolling averages and lag features, which represent short-term and long-term price motions, respectively.
- Demand Indicators: Indicators, such as average rating and reviews, have been extracted from user reviews and sales ranks, and hence can be considered as proxies for demand.
- Time-Based Features: Temporal attributes (such as month, day of the week and time of year, e.g., in holiday season) are useful in dealing with seasonal effects on price.

B. Model Selection

The choice of a suitable model is of course influenced by data characteristics and the requirements of the problem. Our goal is to develop a model that has good accuracy, interpretability, and computational efficiency.

1) Linear Regression: It used as a reference model because of its simplicity and interpretability. Although useful for linearisms, it does not always perform well with non-linear data, i.e., complex data.

2) Random Forest: Ensemble learning approach that integrates several decision trees to increase the prediction accuracy. Random Forest is selected because it is very resistant to overfitting and is capable of providing non-linear interaction within data.

3) Support Vector Machines (SVM): Effective for datasets with a high degree of dimensionality. Nevertheless, due to the increased computational expense, SVM was applied to a smaller sample to make a comparison.

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

4) Long Short-Term Memory (LSTM): As a kind of recurrent neural network (RNN), LSTM benefits from its suitability for time-series data and sequences, enabling the model to capture temporal dependency and trend efficiently.

C. Model Training and Evaluation

All models are thoroughly trained and tested to achieve the best performance.

1) Training Process: Models are trained on 80% of the data and the remaining 20% are held out for testing. Hyperparameter tuning is implemented with methods (e.g., grid search and random search), tuning parameters (e.g., tree depth, for Random Forest and learning rate, for LSTM).

Model	MAE	MSE	RMSE
Linear Regression	12.7	24.3	4.9
Random Forest	10.3	19.8	4.4
SVM	11.9	21.4	4.6
LSTM	9.8	18.5	4.3

2) Evaluation Metrics: Model performance is evaluated through Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), and RMSE is selected as the main evaluation metric as it penalizes large errors.

According to the results, LSTM gets the smallest RMSE and clearly seems to offer an advantage in the time-series nature of the data. Random Forest has good performance as well, compromising accuracy for computational efficiency, and thus can be deployed in production.

VII. COMPARATIVE ANALYSIS

A. Comparison Framework

In order to find the best model, we empirically compare the performance of all models on the test set, with respect to accuracy, scalability and interpretability. LSTM and Random Forest emerged as top performers, with LSTM showing slightly better accuracy due to its ability to capture complex temporal relationships. On the other hand, Random Forest has under its favor both interpretability and lower computational cost, both of which are useful for real-time systems.

B. Comparison Table

Feature	Linear Regression	Random Forest	SVM	LSTM
Accuracy	Moderate	High	Moderate	High
Interpretability	High	Moderate	Low	Low
Computation Requirements	Low	Moderate	High	High
Suitability for Real-Time	Good	Excellent	Fair	Good
Handling Non-linear Data	Poor	Excellent	Moderate	Excellent

Based on this comparison, Random Forest is the chosen model for real-time implementation because of its robustness, interpretability, and high computational efficiency. LSTM, good accuracy, but it has high computational costs and could be better suited for batch processing.

VIII. DISCUSSION, LIMITATION AND FUTURE WORK

A. Discussion

The combination of web scraping and machine learning for price prediction has great potential in the e-commerce environment. Prediction of price movements is useful to consumers who can make good buying choices, and to companies who can set good pricing policies. According to our analysis Random Forest and LSTM models demonstrate good prediction reliability, LSTM demonstrating superior accuracy and Random Forest illustrating better computational efficiency. The results point to the fact that a hybrid model based on combined functionalities of both algorithms could lead to further enhancement of prediction accuracy.

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

B. Limitations

Despite its contributions, the study has limitations:

1) Data Source Restrictions: The use of web-scraping imposes a limit in terms of data accessibility, since it relies on information that is publicly available and may not include all those data that are relevant to the price of the product (e.g., data on inventory levels or promotional data).
2) Model Complexity: Such models, LSTM, even accurate, are high computational and thus can act as an impediment in the real-time application. There is still some work needed to accommodate accuracy with computation expense.
3) Ethical Constraints: Ethical constraints can reduce the number of times and areas in which a webpage can be scraped, which might lead to outdated information and inaccuracies in highly dynamic market settings.

C. Future Work

Future research can build upon the current study in many aspects:

1) Hybrid Model Development: Due to the advantages of both Random Forest and LSTM, it is expected to obtain an accurate model with moderate computational cost, which will help to achieve real-time performance.
2) Incorporation of Additional Data Sources: Including complementary data sources (social media trend or stock availability), for example, may enhance model's predictive performance.
3) Real-Time Deployment: The need for the development of a streamlined real-time system for e-commerce prediction, with model optimization techniques in order to lower the latency feasible, would allow practical use in live e-commerce environments.

IX. COCLUSION

In order to find the optimal price, one must find a way through a wide range of options. Methodology using a combination of Random Forest and Long Short-Term Memory (LSTM) models is used for the prediction of the future values of the price, in light of past trends and other intervening variables.

Our results show that these machine learning models have their own advantages. Random Forest model is very effective in the case of structured data such as table with only a minimum amount of preprocessing in predicting the price trend of short term. On the other hand, the LSTM model, thanks to its capacity to model temporal relationships, is more appropriate for long-term predictions and seasonal patterns in pricing data. Through this dual-model paradigm there is flexibility, enabling users to apply insights according to their own needs, e.g., price changes or purchase fore-casting.

While the tool demonstrates promising results, certain limitations remain. These are, for instance, the requirement of continuous data refresh for achieving good prediction accuracy, and the possibility of inconsistent results arising from data heterogeneity across different platforms. Furthermore, real-time data processing - related computational costs - may limit the scalability of larger deployments.

Although limited by such constraints, our work highlights the valuable role that predictive analytics can play in e-commerce. To the consumer, it is a promise of better decision-making and cost-neutrality. For companies it provides information on pricing trends and competitive standing. Future work will be dedicated towards increasing the accuracy of the models, decreasing the computational cost, and investigating the implementation in a variety of other e-commerce tools to offer an even smoother experience. In summary, this study performs a significant application of machine learning and web scraping, highlighting the importance of machine learning and web scraping in the digital economy of the day.

REFERENCES

[1] D. Shaikh, R. Khan, K. Panokher, M. Kr. Ranjan, and V. Sonaje, "E-commerce Price Comparison Website using Web Scraping," International journal of innovative research in engineering & multidisciplinary physical sciences, Jun. 25, 2023. Available: https://doi.org/10.37082/ijirmps.v11.i3.230223
[2] S. Rajendar, K. Manikanta, M. Mahendar, and K. Madhavi, "Price Comparison Website for Online Shopping," IJCRT, Jun.2021, [Online]. Available: https://ijcrt.org/papers/IJCRT2106450.pdf
[3] H. Khatter, Dravid, A. Sharma, and A. Kumar Kushwaha, "Web Scraping based Product Comparison Model for E- Commerce Websites," IEEE Conference Publication | IEEE Xplore, Jul. 29, 2022 . Available:https://ieeexplore.ieee.org/document/991582
[4] Martina D'Souza, Soham Desai, "Web Scraping based Product Comparison Model for E-Commerce Website" IEEE Conference Publication | IEEE Xplore, April 2024. Available:https://www.jetir.org/papers/JETIR2404G32.pdf?form=MG0AV3

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

[5] Faizan Raza Sheikh, Hrishabh V. Petkar, Abdul Malik Sheikh, Harsh Lalchand Kose, A. Chandekar, and P. V. Awale, "Price Comparison using Web-scraping and Data Analysis," International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), Jun. 2023, [Online]. Available: https://ijarsct.co.in/Paper11354.pdf

[6] D. G. Madhusudhan, N. G. Bhat, 2Sahana Venkatraman Patgar, Chandan N A, and Bharath S V, "E-COMMERCE PRODUCT PRICE TRACKER," JETIR, Jun. 2021, [Online].
Available:https://www.jetir.org/papers/JETIR2106691.pdf

[7] M. SOWMIYA, S. CS, M. RAJA M, and S. KUMAR S,"Price Comparison for Products in Various E-Commerce Website," IJRTI, 2023, [Online].
Available: https://www.ijrti.org/papers/IJRTI2305089

[8] Shalini and R. Ambikapathy, "E-Commerce Analysis and Product Price Comparison Using Web Mining," International Journal of Research Publication and Reviews, Jun. 2022, [Online].
Available:https://ijrpr.com/uploads/V3ISSUE6/IJRPR5223.pdf

[9] S. Shreekumar, S. Mundke, and M.Dhanawade, "IMPORTANCE OF WEB SCRAPING IN E- COMMERCE BUSINESS," NCRD's Technical Review : e-Journal, Jan. 2022,[Online].
Available:https://www.ncrdsims.edu.in/ckfinder/userfiles/files/04(2).pdf

[10] K. Varun, P. Rajesh, P. Dileep, and B.S.V.Satish, "Price Comparison for Online Shopping," International Journal of Innovative Science and Research Technology, Dec. 2023, [Online].
Available:https://www.ijisrt.com/assets/upload/files/IJISRT23DEC258.pdf

[11] S. Mehak, R. Zafar, S. Aslam, and S. Masood Bhatti, "Exploiting Filtering approach with Web Scrapping for Smart Online Shopping," International Conference on Computing, Mathematics and Engineering Technologies, 2019,[Online].
Available: https://ieeexplore.ieee.org/document/86733

[12] Ben Powell, Guy Nason, Duncan Elliott, Matthew Mayhew, Jennifer Davies, Joe Winton, Tracking and Modelling Prices Using Web-Scraped Price Microdata: Towards Automated Daily Consumer Price Index Forecasting, Journal of the Royal Statistical Society Series A: Statistics in Society, Volume 181, Issue 3, June 2018, Pages 737–756.
Available: https://doi.org/10.1111/rssa.12314

[13] G. V. Kumar, "Web Scraping for E-Commerce Websites," International Journal for Research in Applied Science and Engineering Technology, vol. 10, no. 6, pp. 4114–4119, Jun. 2022.
Available: https://doi.org/10.22214/ijraset.2022.44841

[14] S. B Peedi, Sangeeta , V. Gaded, and Vedhika, "A Novel of Product Price Analysis Using Web Scraping," IJIRT, Aug. 2022, [Online].
Available:https://ijirt.org/master/publishedpaper/IJIRT156159_PAPER.pdf

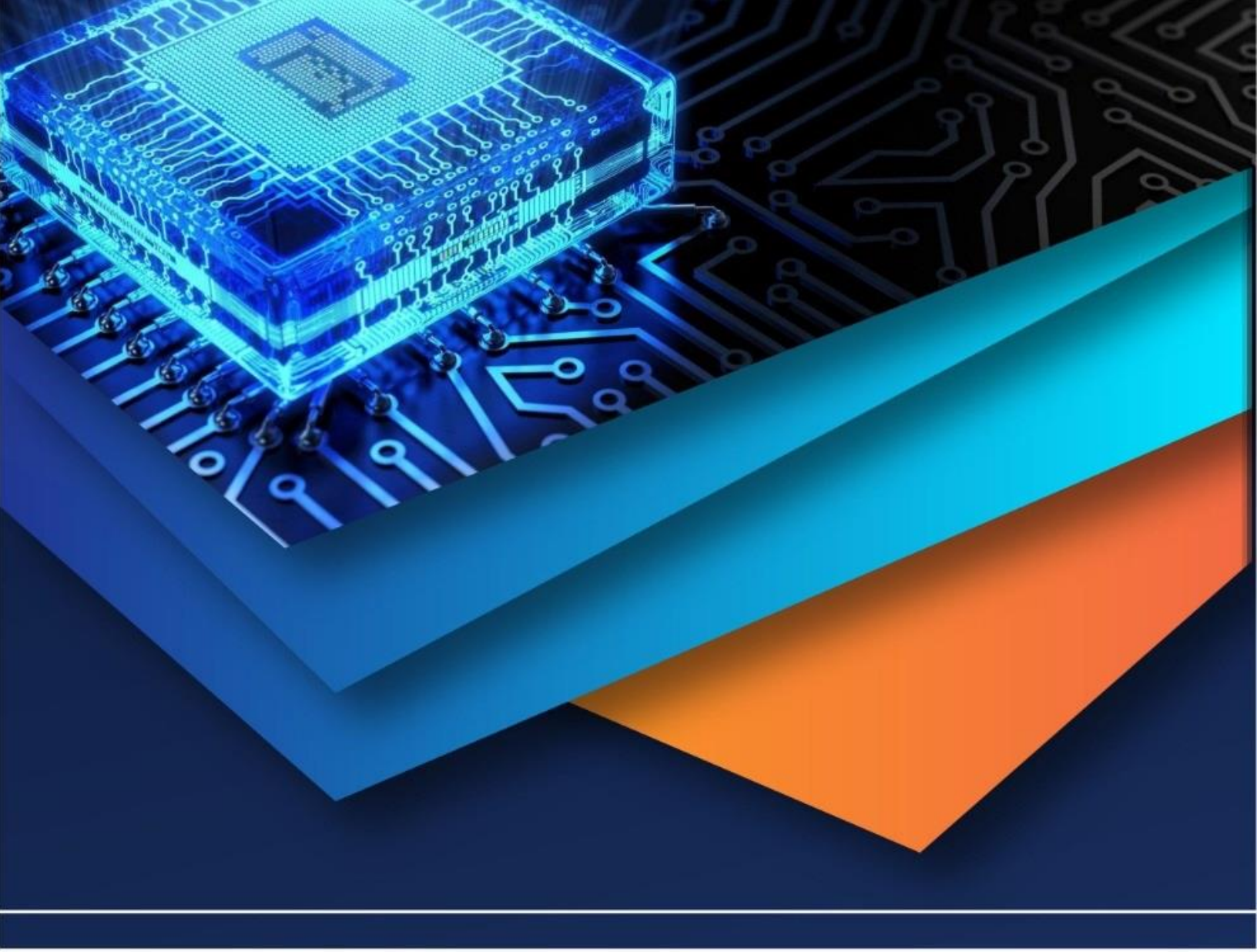
[15] Ambre, P. Gaikwad, K. Pawar, and V. Patil, "Web and Android Application for Comparison of E-Commerce Products," International Journal of Advanced Engineering, Management and Science, vol. 5, no. 4, pp. 266–268, Jan. 2019,
Available: https://doi.org/10.22161/ijaems.2019.545

[16] Pavan Sai Rayalla, Setty Vasu Vivek, Sreeja G, Sneha Sanju, and Dr. S. Vijay Kumar, "PRODUCT PRICE COMPARISON ON MULTIPLE E-COMMERCE WEBSITES," International Research Journal of Modernization in Engineering Technology and Science, Jun.2023,
Available: https://doi.org/10.56726/irjmets.2023.40965

[17] Vaibhavi B Raj, A S Sushmitha Urs, Abhishek Kumar Pandey, Jagruthi G, Archana VR, Deepthi Das "Price Probe: E-Commerce Platforms Using Machine Learning" International Journal for Multidisciplinary Research (IJFMR) E-ISSN: 2582-2160
Available:https://www.ijfmr.com/papers/2024/3/20185.pdf?form=MG0AV3

[18] D. K. Sharma, S. Lohana, S. Arora, A. Dixit, M. Tiwari, and T. Tiwari, "E-Commerce product comparison portal for classification of customer data based on data mining," Materials Today: Proceedings, vol. 51, pp. 166–171, 2022
Available: https://doi.org/10.1016/j.matpr.2021.05068

©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)