

# SMART SHOPPERS

A Project Report

Submitted By

**P.Harshini**

**210304124219**

**P.Shashitha**

**210304124204**

**V.Chetana**

**210304124415**

**V.Manoj**

**210304124426**

in Partial Fulfilment For the Award of

the Degree of

BACHELOR OF TECHNOLOGY

COMPUTER SCIENCE & ENGINEERING

Under the Guidance of

**Prof. Ritu Agrawal, Prof. Gourav Yadav**

Assistant Professor



VADODARA

October - 2024



# PARUL UNIVERSITY

## CERTIFICATE

This is to Certify that Project - II (203105400) of 7<sup>th</sup> Semester entitled “**SMART SHOPPERS**” of Group No. **PUCSE\_114** has been successfully completed by

- P.HARSHINI - 210304124219
- P.SHASHITHA - 210304124204
- V.CHETANA - 210304124415
- V.MANOJ - 210304124426

under my guidance in partial fulfillment of the Bachelor of Technology (B.Tech) in Computer Science & Engineering of Parul University in Academic Year 2023- 2024.

Date of Submission :\_\_\_\_\_

**Prof. Ritu Agrawa , Prof. Gourav Yadav**

Project Guide

**Dr. Amit Barve,**

Head of Department,

CSE, PIET,

Project Coordinator:-

**Dr.Kurti Sutaria**

Parul University.

## **ACKNOWLEDGEMENT**

*“The only way to do great work is to love what you do.”*

-Steve-Jobs

We extend our sincere gratitude to everyone who played a pivotal role in the successful completion of the "Smart Shopper" project. Firstly, we are deeply thankful to our project supervisors, **Prof. Ritu Agrawal** and **Prof. Gourav Yadav**, for their unwavering support, insightful guidance, and continuous encouragement throughout the project's development. Their expertise significantly influenced the project's direction, contributing to its successful outcome. We also wish to express our appreciation to the faculty members of the **CSE Department** and our esteemed Head of Department, **Prof. Amit Brave sir**, for their ongoing support and constructive feedback, which greatly enhanced the quality of our project.

A heartfelt thanks to our team members—Harshini, Shashitha, Chetana, and Manoj—for their dedication, collaboration, and hard work in making this project a reality. Their commitment to teamwork and excellence was instrumental in overcoming the challenges we faced and achieving our objectives.

We also extend our gratitude to the users and testers who provided invaluable feedback and suggestions that helped us refine the project, making it more functional and user-friendly. Additionally, we acknowledge the creators of the libraries, frameworks, and tools utilized during the project. Their innovative contributions played an essential role in the project's implementation.

Lastly, we are profoundly grateful to our classmates, friends, and families for their steadfast support, encouragement, and patience throughout the duration of this project. With deep appreciation, we recognize the invaluable contributions of all individuals and entities mentioned above, without whom the successful completion of this project would not have been possible.

**P.Harshini-210304124219**

**P.Shashitha-210304123204**

**V.Chetana-210204124415**

**V.Manoj-210304124426**

**CSE, PIET**

**Parul University,**

**Vadodara**

## ABSTRACT

The "Smart Shopper" project aims to streamline online shopping by offering users a tool to track product prices across multiple e-commerce platforms, simplifying their purchasing decisions. This report outlines the project's objectives, development process, and potential future advancements.

As e-commerce continues to grow, shoppers are often overwhelmed by the sheer number of product options available. In response to this, "Smart Shopper" was created to help users navigate the online shopping landscape more efficiently. This web-based application employs cutting-edge web scraping technology to gather real-time product information from various e-commerce websites. It allows users to filter and rank products based on key criteria such as price, brand, product features, and customer reviews, making it easier for them to find items that suit their needs and budgets.

Beyond basic price tracking, "Smart Shopper" integrates machine learning algorithms to analyze user behavior and preferences. This enables the app to provide personalized product recommendations, enhancing the shopping experience for each individual. By leveraging data on past purchases and user interactions, the application tailors its suggestions to fit unique preferences, fostering greater customer satisfaction and loyalty.

In addition to serving consumers, "Smart Shopper" offers valuable insights to retailers and e-commerce platforms. By analyzing user search patterns, purchase history, and product preferences, the app helps businesses understand consumer behavior and market trends. This data allows retailers to fine-tune their product offerings, pricing strategies, and marketing campaigns, ultimately boosting sales and competitiveness in the e-commerce sector.

Overall, "Smart Shopper" represents a significant step forward in enhancing both customer experience and retailer capabilities. By simplifying the purchasing process and providing actionable insights to sellers, the app has the potential to shape the future of e-commerce and transform how we shop online.

**Key Points:** **web scraping, e-commerce, price tracking, price comparison.**

# Table of Contents

<b>Acknowledgements</b>	iii
<b>Abstract</b>	iv
<b>List of Tables</b>	ix
<b>List of Figures</b>	xi
<b>1 INTRODUCTION</b>	1
1.1 E-commerce Environment Analysis . . . . .	1
1.2 Purpose and Scope of the Project . . . . .	1
1.3 Technical Framework . . . . .	2
1.4 User-Centered Design Model . . . . .	2
1.5 Value for Customers and Informed Purchasing Decisions . . . . .	3
<b>2 LITERATURE SURVEY</b>	4
2.1 Web Scraping for E-Commerce Websites . . . . .	4
2.2 Price Dynamics in E-commerce: A web scraping study . . . . .	5
2.3 E-Commerce Price Comparison Website Using Web Scraping . . . . .	6
2.4 Product comparison website using web scraping and machine learning . . . . .	7
2.5 Scraping and Visualization of Product Data from E-commerce Website . . . . .	8
2.6 Importance of Web Scraping in E-commerce and E-marketing . . . . .	9

2.7	Importance of Web Scraping in E-commerce Business . . . . .	10
2.8	Using Web Scraping In A Knowledge Environment To Build Ontologies Using Python And Scrapy . . . . .	11
2.9	Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application	12
2.10	Increasing Online Shop Revenues with Web Scraping: A Case Study for the Wine Sector . . . . .	13
2.11	Web Scraping Techniques to Collect Data on Consumer Electronics and Airfares for Italian HICP Compilation . . . . .	14
2.12	Web Scraper Revealing Trends of Target Products and New Insights in Online Shopping Websites . . . . .	15
2.13	Commodity Price Data Analysis Using Web Scraping . . . . .	16
2.14	Legality and Ethics of Web Scraping . . . . .	17
2.15	Recommendation System Using Product Rank Algorithm For E-Commerce . . . . .	18
2.16	A Web Scraping Framework for Descriptive Analysis of Meteorological Big Data for Decision-Making Purposes . . . . .	19
2.17	Implementation of Web Scraping for Journal Data Collection on the SINTA Website	20
2.18	An Intelligent Survey of Personalized Information Retrieval using Web Scraper . .	21
2.19	News Aggregation using Web Scraping News Portals . . . . .	22
2.20	Forecasting Prices of Fish and Vegetable using Web Scrapped Price Micro Data . .	23
<b>3</b>	<b>ANALYSIS / SOFTWARE REQUIREMENTS SPECIFICATION (SRS)</b>	<b>24</b>
3.1	Introduction . . . . .	24
3.1.1	Purpose . . . . .	24
3.1.2	Terms . . . . .	24
3.1.3	Intended Audience and Readers . . . . .	24
3.1.4	Product Scope . . . . .	24
3.1.5	References . . . . .	25

3.2 General . . . . .	25
3.2.1 Products . . . . .	25
3.2.2 Product Features . . . . .	25
3.2.3 User groups and characteristics . . . . .	26
3.2.4 Operating environment . . . . .	26
3.2.5 Design and Functionality . . . . .	26
3.2.6 User Information . . . . .	27
3.2.7 Assumptions and Dependencies . . . . .	27
3.3 External Interface Requirements . . . . .	27
3.3.1 User Interface . . . . .	27
3.3.2 Hardware Interface . . . . .	28
3.3.3 Software Interface . . . . .	28
3.3.4 Communication . . . . .	28
3.4 Functions and requirements . . . . .	28
3.4.1 Functional requirements . . . . .	28
<b>4 SYSTEM DESIGN</b>	<b>31</b>
4.1 System Architecture . . . . .	31
4.2 User Interface Design . . . . .	33
4.3 Database Design . . . . .	35
<b>5 METHODOLOGY</b>	<b>37</b>
5.1 Technology Stack . . . . .	37
5.2 Development Process . . . . .	37
5.3 Testing and Evaluation . . . . .	39
<b>6 IMPLEMENTATION</b>	<b>40</b>
6.1 Introduction . . . . .	40

6.2	Technology Stack . . . . .	40
6.3	Front End Development . . . . .	40
6.4	Back End Development . . . . .	40
6.5	Web Scraping Implementation . . . . .	43
6.6	Data Processing and Insights . . . . .	43
6.7	Challenges and Solutions . . . . .	43
<b>7</b>	<b>TESTING</b>	<b>44</b>
7.1	Introduction . . . . .	44
7.2	Unit Testing . . . . .	44
7.3	Integration Testing . . . . .	44
7.4	User Acceptance Testing . . . . .	46
7.5	Test Cases and Results . . . . .	46
7.6	Performance Testing . . . . .	47
7.7	Bug Tracking and Resolution . . . . .	47
7.8	Final Testing and Validation . . . . .	47
<b>8</b>	<b>CONCLUSION</b>	<b>49</b>
8.1	Summary of Results . . . . .	49
8.2	Key Findings . . . . .	49
8.3	Challenges . . . . .	49
8.4	Lessons Learned . . . . .	49
8.5	Future Directions . . . . .	50
8.6	Conclusion . . . . .	50
<b>9</b>	<b>FUTURE WORK</b>	<b>51</b>

# List of Tables

2.1	Web Scraping for E-Commerce Websites . . . . .	4
2.2	Price Dynamics in E-commerce: A web scraping study . . . . .	5
2.3	E-Commerce Price Comparison Website Using Web Scraping . . . . .	6
2.4	Product comparison website using web scraping and machine learning . . . . .	7
2.5	Scraping and Visualization of Product Data from E-commerce Website . . . . .	8
2.6	Importance of Web Scraping in E-commerce and E-marketing . . . . .	9
2.7	Importance of Web Scraping in E-commerce Business . . . . .	10
2.8	Using Web Scraping In A Knowledge Environment To Build Ontologies Using Python And Scrapy . . . . .	11
2.9	Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application	12
2.10	Increasing Online Shop Revenues with Web Scraping: A Case Study for the Wine Sector . . . . .	13
2.11	Web Scraping Techniques to Collect Data on Consumer Electronics and Airfares for Italian HICP Compilation . . . . .	14
2.12	Web Scraper Revealing Trends of Target Products and New Insights in Online Shopping Websites . . . . .	15
2.13	Commodity Price Data Analysis Using Web Scraping . . . . .	16
2.14	Legality and Ethics of Web Scraping . . . . .	17
2.15	Recommendation System Using Product Rank Algorithm For E-Commerce . . . .	18
2.16	A Web Scraping Framework for Descriptive Analysis of Meteorological Big Data for Decision-Making Purposes . . . . .	19
2.17	Implementation of Web Scraping for Journal Data Collection on the SINTA Website	20
2.18	An Intelligent Survey of Personalized Information Retrieval using Web Scraper . .	21
2.19	News Aggregation using Web Scraping News Portals . . . . .	22
2.20	Forecasting Prices of Fish and Vegetable using Web Scrapped Price Micro Data . .	23

4.1 Database Schema . . . . .	35
-------------------------------	----

# List of Figures

4.1	UML Diagram	32
4.2	Sequence Diagram	33
4.3	DFD Level-0	34
4.4	DFD Level-1	34
4.5	ER Model	36
4.6	Activity Diagram	36
6.1	Home Page 1	41
6.2	Home Page 2	41
6.3	Frontend 1	42
6.4	Frontend 2	42
7.1	Unit Testing of Web Scraping Functions	45
7.2	Unit Testing of API Endpoints	45
7.3	Testing 3	47
7.4	Testing 4	48

# **Chapter 1**

## **INTRODUCTION**

### **1.1 E-commerce Environment Analysis**

The rapid growth of e-commerce has changed the way people shop. Today, consumers have access to a variety of products and services from various online sites at their fingertips. This change makes it easier for customers to search, compare, and purchase. However, too many options can lead to decision fatigue, which leaves consumers feeling overwhelmed by the many options available. Additionally, varying prices, varying product quality, and a plethora of user reviews often make it difficult for users to make a purchasing decision. [3].

To solve these problems, the Smart Shopper program offers new solutions that make online shopping easier. The application uses web scraping technology and advanced algorithms to provide users with personalized recommendations, allowing them to find the products that best suit their interests. This not only saves time, but also increases overall satisfaction, as customers can easily find what they are looking for without being distracted by irrelevant options.

### **1.2 Purpose and Scope of the Project**

The “Smart Shopper” project aims to bridge the gap between consumers and their ideal products by building a powerful web scraping tool. The main goal is to create an application that can collect and analyze product data from various e-commerce sites and then display it in the user interface for easy comparison and selection.

The scope of this project goes beyond simple cost tracking. In addition to collecting product price information, the application will also analyze other important factors such as product availability, customer reviews, sales ratings, and reputation. By doing this, the application can provide users with an overview of the available options, helping them make informed decisions based on various factors. [6].

The project also aims to implement real-time data updates, ensuring that users have access to the most current information. Additionally, machine learning models will be incorporated to refine and enhance the recommendation system, offering users personalized suggestions that become more accurate over time.

### 1.3 Technical Framework

The technical foundation of the "**Smart Shopper**" project is built on a robust architecture that integrates web scraping, data processing, and machine learning technologies. The web scraping component is responsible for collecting data from various e-commerce websites. This involves navigating through product pages, extracting key details such as prices, descriptions, reviews, and images, and storing them in a structured format.

To manage large volumes of data efficiently, the project employs advanced data processing tools that filter and organize the extracted information. This allows the system to deliver relevant results to users, while also ensuring that the data is updated frequently. Furthermore, sophisticated algorithms are utilized to rank and sort products based on user preferences, such as price range, brand affinity, or specific product features. [3].

At the core of the recommendation system is a machine learning framework that continuously learns from user interactions. As users browse products and make purchases, the system gathers data on their behavior, such as search patterns, preferences, and feedback. This data is then used to train the model, enabling it to provide more accurate and personalized recommendations over time. The use of artificial intelligence and predictive analytics enhances the overall user experience, making the shopping process more efficient and enjoyable.

### 1.4 User-Centered Design Model

The success of the "**Smart Shopper**" application hinges on its user-centered design, which prioritizes usability, functionality, and satisfaction. The design process began with an in-depth analysis of user needs and preferences, ensuring that the interface is intuitive and easy to navigate. A key focus of the design model is to minimize the complexity of the shopping experience while maximizing user engagement.

The application features a clean and modern interface, with simple navigation tools that allow users to browse through products, apply filters, and sort results based on various criteria. Interactive elements, such as product comparison tools and user-generated review summaries, enhance the shopping experience by giving users a clear overview of the options available to them. Additionally,

the design incorporates responsive elements, ensuring that the application works seamlessly across different devices, whether on a desktop or a mobile phone. [4].

User feedback is another integral aspect of the design model. The application includes features that allow users to provide ratings and reviews, which in turn are used to refine the product recommendations and improve the overall functionality of the system. By incorporating feedback loops into the development process, the app is continuously evolving to better meet the needs of its users.

## **1.5 Value for Customers and Informed Purchasing Decisions**

The "Smart Shopper" application provides significant value for both consumers and retailers alike. For consumers, the app not only saves time by filtering and recommending products based on their specific needs, but it also helps them make informed purchasing decisions by presenting critical information such as price comparisons, product reviews, and seller ratings.

The cost-saving benefits of the application are particularly noteworthy. By providing real-time price updates and product comparisons, users can easily identify deals and discounts, ultimately leading to more cost-effective purchases. Furthermore, the personalized recommendations generated by the machine learning algorithms ensure that users are presented with relevant options, reducing the likelihood of impulse buys or purchasing products that do not meet their expectations.

For retailers and e-commerce platforms, the "**Smart Shopper**" application offers valuable insights into customer behavior and market trends. By analyzing user preferences, purchasing habits, and search patterns, the app generates analytics reports that help businesses optimize their product offerings, pricing strategies, and marketing campaigns. These insights enable retailers to stay competitive in the crowded e-commerce landscape by tailoring their offerings to meet the demands of their target audience.

In conclusion, the "Smart Shopper" project represents a significant advancement in the e-commerce sector, combining the latest web scraping technologies with machine learning and user-centered design. By providing value to both consumers and businesses, the application has the potential to revolutionize the way people shop online, making the experience more efficient, personalized, and enjoyable. [7].

# **Chapter 2**

## **LITERATURE SURVEY**

### **2.1 Web Scraping for E-Commerce Websites**

Student Name:	Pothineni Harshini		
Enrollment No:	210304124219	Branch:	CSE
Title Of Jounral Paper:	Web Scraping for E-Commerce Websites		
Authors:	Gandhe Vineeth Kumar, Hema M S, Aishwarya R, K R Mamatha		
Journal/Conference:	Journal of emerging technologies and innovative research		
Volume/Issue:	2022	Pages:	11

Table 2.1: Web Scraping for E-Commerce Websites

Abstract: Web scraping has emerged as a crucial device for extracting precious facts from e-commerce websites, supporting organizations benefit insights into merchandise, pricing, and patron behavior. This case examine offers an outline of the internet scraping techniques hired in the e-commerce region, examining strategies used to collect product info, pricing data, consumer critiques, and different applicable records from on line stores. The dialogue additionally addresses challenges, moral concerns, and felony aspects related to web scraping, along with first-rate practices for making sure responsible and compliant use of this era within the e-commerce area. Through a comprehensive overview of current studies and literature, this newsletter targets to offer researchers and enterprise specialists with a higher know-how of both the advantages and barriers of net scraping in e-commerce, as well as practical pointers for successful implementation. [3].

## 2.2 Price Dynamics in E-commerce: A web scraping study

Student Name:	Pothineni Harshini		
Enrollment No:	210304124219	Branch:	CSE
Title Of Jounral Paper:	Price Dynamics in E-commerce: A web scraping study		
Authors:	Arman Shaikh Raihan Khan, Komal Panokher, Mritunjay Kumar Ranjan		
Journal/Conference:	Journal of information systems(JIS)		
Volume/Issue:	2022	Pages:	6

Table 2.2: Price Dynamics in E-commerce: A web scraping study

**Abstract:** This research paper explores the understanding of e-commerce value through the use of web scraping technology. It aims to analyze the pricing behaviors of different online retailers by applying various web scraping tools and techniques. The study delves into factors that affect pricing decisions, such as market demand, competitive pricing strategies, and the stages of a product's life cycle. Through analyzing data gathered from web scraping, the research identifies important patterns and trends in e-commerce pricing dynamics. It further examines how factors like price flexibility, consumer preferences, and market competition influence the pricing approaches adopted by e-commerce platforms. The study's findings offer valuable insights into the complexities of pricing within the digital economy, aiding in cost optimization and strategic decision-making in the e-commerce sector. It also emphasizes the role of web scraping as an effective method for detecting price changes and supporting marketing strategies. Ultimately, this research highlights the critical role of data-driven insights in achieving a competitive advantage and fostering growth in the fast-changing e-commerce environment. [7].

## 2.3 E-Commerce Price Comparison Website Using Web Scraping

Student Name:	Pothineni Harshini		
Enrollment No:	210304124219	Branch:	CSE
Title Of Journal Paper:	E-Commerce Price Comparison Website Using Web Scraping		
Authors:	Arman, Raihan, Komal, Mritunjay, Vaibhav		
Journal/Conference:	IJIR		
Volume/Issue:	2023	Pages:	10

Table 2.3: E-Commerce Price Comparison Website Using Web Scraping

**Abstract:** This research paper discusses the creation and deployment of an e-commerce price comparison website utilizing web scraping technology. It outlines the design and architecture of the system, which uses web scraping tools to collect product details, pricing data, and other relevant information from multiple online retail platforms. The website enables users to gain insights into product availability, price fluctuations, and discounts across different e-commerce sites through data comparison. The paper delves into the technical aspects of web scraping, including data extraction, analysis, and storage, while addressing challenges such as adapting to changes in website structures and maintaining data accuracy. Furthermore, the study evaluates the website's usability and effectiveness through user testing and feedback. The results suggest that web scraping can enrich the e-commerce experience by helping consumers make informed purchasing decisions and promoting competition within the online retail industry. [1].

## 2.4 Product comparison website using web scraping and machine learning

Student Name:	Pothineni Harshini		
Enrollment No:	210304124219	Branch:	CSE
Title Of Journal Paper:	Product comparison website using web scraping and machine learning		
Authors:			
Journal/Conference:	IEEE		
Volume/Issue:	2022	Pages:	9

Table 2.4: Product comparison website using web scraping and machine learning

**Abstract:** This research paper introduces an innovative method for product comparison by integrating web scraping with machine learning techniques. It describes the development of a product comparison website that automatically collects data from multiple e-commerce platforms through web scraping. The website goes beyond basic data collection, capturing detailed information such as customer reviews, ratings, and product specifications to provide comprehensive comparisons. The study also details the application of machine learning algorithms to enhance the analysis of the collected data.

Additionally, the paper evaluates the website's performance and accuracy through thorough testing and validation using real-world data. The findings demonstrate that combining web scraping with machine learning can significantly improve the effectiveness of product comparison tools. These tools deliver in-depth, relevant information to consumers, supporting them in making well-informed purchasing choices within the competitive landscape of modern e-commerce. [8].

## 2.5 Scraping and Visualization of Product Data from E-commerce Website

Student Name:	Pothineni Harshini		
Enrollment No:	210304124219	Branch:	CSE
Title Of Journal Paper:	Scraping and Visualization of Product Data from E-commerce Website		
Authors:	V. Srividhya, P. Megala		
Journal/Conference:	International Journal of Computer Sciences and Engineering		
Volume/Issue:	2019	Pages:	9

Table 2.5: Scraping and Visualization of Product Data from E-commerce Website

**Abstract:** This research paper provides a comprehensive analysis of collecting product data from e-commerce websites and visualizing it for analytical insights. It covers various web scraping techniques, including methods for data extraction, analysis, and storage. The study reviews widely used web scraping tools and libraries, such as Beautiful Soup and Scrapy, discussing their advantages and limitations. It also addresses common challenges in the scraping process, such as managing dynamic content, navigating complex website structures, and ensuring data quality.

Additionally, the paper delves into visualizing the scraped data, focusing on methods and tools that enhance data accessibility and insights. It showcases the use of popular visualization libraries like Matplotlib, Seaborn, and Plotly to create visualizations, including histograms, scatterplots, and heatmaps. These visual representations are used to analyze aspects such as product features, pricing patterns, and customer feedback, providing a clearer understanding of the data. [13].

## 2.6 Importance of Web Scraping in E-commerce and E-marketing

Student Name:	Pitchuka Shashitha			
Enrollment No:	210304124204	Branch:	CSE	
Title Of Jounral Paper:	I	Importance of Web Scraping in E-commerce and E-marketing		
Authors:	Kasereka Henrys			
Journal/Conference:	HK Cooperation			
Volume/Issue:	2021	Pages:	15	

Table 2.6: Importance of Web Scraping in E-commerce and E-marketing

**Abstract:** This research paper underscores the pivotal role of web scraping in enhancing success and competitiveness within the e-commerce sector. It examines the benefits of using web scraping techniques to extract valuable data from online platforms, offering businesses actionable insights and strategic advantages. Through a detailed analysis of the e-commerce landscape, the paper illustrates how web scraping facilitates the collection of critical information, such as product details, pricing trends, customer feedback, and competitor strategies.

The use of web scraping technology allows businesses to gain a deeper understanding of market trends, consumer behavior, and competitive dynamics. The study also explores the strategic significance of web scraping in e-commerce, highlighting its role in refining pricing strategies, broadening product portfolios, and discovering new market opportunities. The results emphasize that web scraping is essential for e-commerce operations, enabling businesses to implement data-driven strategies that foster sustainable growth and maintain a competitive edge in the digital marketplace. [10].

## 2.7 Importance of Web Scraping in E-commerce Business

Student Name:	Pitchuka Shashitha		
Enrollment No:	210304124204	Branch:	CSE
Title Of Jounral Paper:	Importance of Web Scraping in E-commerce Business		
Authors:	Sandeep Shreekumar, Satyan Mundke, Dr. Murlidhar Dhanawade		
Journal/Conference:	Association of Computer Machinery (ACM)		
Volume/Issue:	2022	Pages:	3

Table 2.7: Importance of Web Scraping in E-commerce Business

**Abstract:** This article investigates the use of web scraping techniques in knowledge environments to facilitate ontology construction. It focuses on leveraging Python and the Scrapy framework to automate the extraction of structured data from a variety of online sources. The study outlines a systematic process for scraping relevant information from the web and converting it into structured formats suitable for ontology development.

Web scraping tools empower researchers to gather large datasets from a range of sources, including scientific literature, online databases, and specialized websites. The collected data is then processed to identify concepts, relationships, and entities that form the core elements of the ontology. This paper emphasizes the advantages of using web scraping in ontology construction, such as access to current and diverse data sources, streamlined data collection, and the ability to quickly prototype ontologies.

By demonstrating the efficacy of web scraping in knowledge acquisition and ontology development, this research supports progress in ontology engineering. It highlights the importance of integrating web scraping into ontology development processes to create more dynamic and robust knowledge structures. [2].

## 2.8 Using Web Scraping In A Knowledge Environment To Build Ontologies Using Python And Scrapy

Student Name:	Pitchuka Shashitha		
Enrollment No:	210304124204	Branch:	CSE
Title Of Jounral Paper:	Using Web Scraping In A Knowledge Environment to Build Ontologies		
Authors:	Hassan Chaib, Krit Salah-ddine, H. Chaib		
Journal/Conference:	EJMCM		
Volume/Issue:	2023	Pages:	9

Table 2.8: Using Web Scraping In A Knowledge Environment To Build Ontologies Using Python And Scrapy

**Abstract:** This research paper provides an in-depth analysis of web scraping and web crawling, examining the latest techniques, methods, and their applications across diverse domains. It clarifies the fundamental differences between web scraping and crawling, highlighting their distinct roles in data extraction and information retrieval from the web. By reviewing current literature and studies, this paper uncovers the complexities associated with web scraping and crawling, such as data extraction strategies, analytical methods, and scalability challenges.

The paper also explores the ethical, legal, and regulatory aspects of these practices, emphasizing the importance of responsible data collection in alignment with privacy standards and copyright laws. Furthermore, it investigates the varied uses of web scraping and crawling, covering areas like e-commerce, social media analysis, academic research, and business intelligence. It demonstrates how these methods can be leveraged to extract meaningful insights, track market dynamics, monitor competitors, and support strategic decision-making.

Overall, this study offers a detailed overview of the current landscape of web scraping and crawling, shedding light on their methodologies, challenges, and practical applications. It serves as a valuable guide for researchers, industry professionals, and organizations aiming to utilize these tools for informed decision-making and advanced data analysis in today's digital world. [9].

## 2.9 Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application

Student Name:	Pitchuka Shashitha		
Enrollment No:	210304124204	Branch:	CSE
Title Of Jounral Paper:	Web Scraping or Web Crawling: State of Art and Application		
Authors:	Moaiad Ahmad Khder		
Journal/Conference:	Al-Zaytoonah University of Jordan (ZUJ)		
Volume/Issue:	2021	Pages:	15

Table 2.9: Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application

**Abstract:** This paper presents an in-depth overview of web scraping and web crawling, exploring the latest technologies, methods, and applications across different fields. It distinguishes between web scraping and web crawling, emphasizing their unique functions in extracting data and information from the web. By thoroughly analyzing current studies and literature, the paper explains the complex processes involved in these practices, including methods for data extraction, parsing, and challenges related to scalability.

Additionally, the study examines ethical, legal, and regulatory issues, emphasizing the importance of conducting data collection responsibly while complying with privacy laws and regulations. It further explores the various applications of web scraping and crawling, highlighting their use in e-commerce, social media analysis, academic research, and business intelligence. The paper demonstrates how these technologies can be applied to gain valuable insights, track market dynamics, monitor job candidates, and inform decision-making.

In summary, this research offers a comprehensive review of the recent progress in web scraping and crawling, detailing the methods, approaches, and practical uses. It serves as a useful resource for researchers, professionals, and organizations aiming to utilize these tools for advanced data analysis and information discovery in the digital era. [12].

## 2.10 Increasing Online Shop Revenues with Web Scraping: A Case Study for the Wine Sector

Student Name:	Pitchuka Shashitha		
Enrollment No:	210304124204	Branch:	CSE
Title Of Jounral Paper:	Increasing Online Shop Revenues with Web Scraping		
Authors:	Joriol, Adria, Josep, Carla, Jordi, Jordi		
Journal/Conference:	British Food Journal (BFG)		
Volume/Issue:	2020	Pages:	9

Table 2.10: Increasing Online Shop Revenues with Web Scraping: A Case Study for the Wine Sector

**Abstract:** This research paper presents a case study on using web scraping techniques to boost revenue for online wine retailers. By analyzing the wine e-commerce landscape, the study showcases how web scraping can be applied to gather valuable data from a variety of online sources. It outlines the methods used to collect information on wine products, pricing trends, customer feedback, and competitor offerings from different e-commerce platforms.

The paper delves into the technical aspects of web scraping, including techniques for data extraction, analysis, storage, and approaches for handling dynamic content while maintaining data accuracy. The analysis of this data reveals key insights into pricing trends, consumer preferences, and market demand patterns in the wine sector.

The findings demonstrate how businesses can use this information to refine pricing strategies, identify new market opportunities, and improve product positioning, ultimately leading to increased revenue for online wine sellers. The study provides concrete evidence of the effectiveness of web scraping in driving revenue growth, emphasizing its role in enabling data-driven decision-making and positioning it as a powerful tool for business optimization and gaining a competitive edge in digital markets. [14].

## 2.11 Web Scraping Techniques to Collect Data on Consumer Electronics and Airfares for Italian HICP Compilation

Student Name:	Varahachalam Chetana		
Enrollment No:	210304124415	Branch:	CSE
Title Of Jounral Paper:	Web Scraping Techniques to Collect Data on Consumer Electronics		
Authors:	Federico Polidoro		
Journal/Conference:			
Volume/Issue:	2015	Pages:	12

Table 2.11: Web Scraping Techniques to Collect Data on Consumer Electronics and Airfares for Italian HICP Compilation

**Abstract:** This paper investigates the results of implementing web scraping techniques for consumer price analysis, specifically targeting the sectors of consumer electronics and airline ticket services. It serves as a foundational report for a study conducted by the Italian National Institute of Statistics (Istat) as part of the European initiative "Multipurpose Price Statistics" (MPS). A central focus of the MPS project is the modernization of data collection through the use of web scraping technologies.

The paper begins with an introduction outlining its main objectives (Section 1) and then describes the criteria for selecting the products tested with web scraping methods (Section 2). Sections 3 and 4 present findings from the analysis of consumer electronics and airline ticket pricing, highlighting both the results obtained and the challenges faced during the application of these techniques. Section 5 discusses the potential improvements in data quality that web scraping may offer in tackling inflation issues. The paper concludes with a summary of key insights in Section 6, emphasizing the implications of big data on statistical practices.

Additionally, two fact boxes are included to highlight crucial aspects of consumer pricing in Italy and the IT solutions employed in the web scraping process. [4].

## 2.12 Web Scraper Revealing Trends of Target Products and New Insights in Online Shopping Websites

Student Name:	Varahachalam Chetana		
Enrollment No:	210304124415	Branch:	CSE
Title Of Jounral Paper:	Web Scraper Revealing Trends of Target Products and New Insights		
Authors:	Habib Ullah, Zahid Ullah, Shahid Maqsood, Abdul Hafeez		
Journal/Conference:	International Journal of Advanced Computer Science and Applications		
Volume/Issue:	2018	Pages:	8

Table 2.12: Web Scraper Revealing Trends of Target Products and New Insights in Online Shopping Websites

**Abstract:** The internet is saturated with a vast array of data, encompassing trillions of Facebook posts, tweets from Twitter, Instagram images, and emails from various servers. This abundance of information creates an urgent need for tools that can efficiently identify frequent updates and extract pertinent details. This research focuses on developing a web scraping tool designed to collect real-time information about specific products from major e-commerce platforms.

The software is built using the Scrapy and Django frameworks and has been configured and tested across several e-commerce websites. Each site produces substantial amounts of product data that need to be scraped. Instead of requiring users to manually browse multiple sites such as amazon.com, alibaba.com, and daraz.pk, the proposed solution offers a unified interface for searching desired products. Additionally, the software features a built-in scheduling function, allowing users to automate data collection at predetermined intervals, ensuring timely access to the information they need. [6].

## 2.13 Commodity Price Data Analysis Using Web Scraping

Student Name:	Varahachalam Chetana		
Enrollment No:	210304124415	Branch:	CSE
Title Of Journal Paper:	Commodity Price Data Analysis Using Web Scraping		
Authors:	Kameswara Rao, Rohit Lagisetty, Maniraj, Dattu, Sneha Ganga		
Journal/Conference:	International Journal of Advances in Applied Sciences (IJAAS)		
Volume/Issue:	2015	Pages:	9

Table 2.13: Commodity Price Data Analysis Using Web Scraping

**Abstract:** Our project centers on the analysis of product price data available online. Analyzing commodity price data is crucial for understanding inflation rates and determining the Consumer Price Index (CPI) in a country. Currently, in some areas, this analysis involves manual data collection from various cities, followed by calculations of inflation and CPI using established formulas. We aim to automate this entire process.

As many consumers increasingly depend on online platforms for their shopping needs, we propose a system that aggregates price data from multiple e-commerce websites for comprehensive analysis. This project introduces a web scraping approach to collect information on various products sold online, process that data, and store it in a centralized database. By automating this process, we eliminate the need for extensive travel and the labor-intensive task of manual data collection. Additionally, the system incorporates web modules that support data analysis and visualization, enhancing the accessibility and usability of the information stored in the database. [5].

## 2.14 Legality and Ethics of Web Scraping

Student Name:	Varahachalam Chetana		
Enrollment No:	210304124415	Branch:	CSE
Title Of Journal Paper:	Legality and Ethics of Web Scraping		
Authors:	Vlad Krotov, Leigh Johnson, Leiser Silva		
Journal/Conference:	Communications of the Association for Information Systems		
Volume/Issue:	2022	Pages:	10

Table 2.14: Legality and Ethics of Web Scraping

Abstract: Automated data extraction from the web, commonly referred to as web scraping, has gained significant traction in both industrial and academic research. Numerous tools and technologies have been developed to facilitate the web scraping process. However, the legal and ethical implications of utilizing these tools for data collection are often overlooked. Neglecting these crucial aspects can lead to ethical dilemmas and potential legal challenges.

This article examines existing legal frameworks along with ethical and privacy considerations, identifying key areas of concern. It also proposes specific questions that researchers and practitioners engaged in web scraping should consider. Addressing these questions can help minimize the risk of ethical and legal conflicts in their work. The article advocates for a balanced approach that allows organizations to utilize web scraping for market analysis and revenue generation while adhering to legal and ethical standards, ensuring responsible data usage. [11].

## 2.15 Recommendation System Using Product Rank Algorithm For E-Commerce

Student Name:	Varahachalam Chetana		
Enrollment No:	210304124415	Branch:	CSE
Title Of Jounral Paper:	Recommendation System Using Product Rank Algorithm For E-Commerce		
Authors:	Aruna Pavate, Urvesh Rathod		
Journal/Conference:	IOSR Journal of Engineering (IOSRJEN)		
Volume/Issue:	2018	Pages:	7

Table 2.15: Recommendation System Using Product Rank Algorithm For E-Commerce

**Abstract:** Automated data extraction from the web, commonly known as web scraping, has become a prevalent practice in both industrial and academic research projects. Numerous tools and technologies have been created to facilitate web scraping initiatives. However, the legal and ethical implications of using these tools for data collection are often overlooked. Ignoring these critical considerations can result in ethical dilemmas and potential legal issues.

This article reviews relevant literature on legal, ethical, and privacy matters to identify key areas of concern. It also presents a series of specific questions that researchers and practitioners involved in web scraping should consider. Addressing these questions can help reduce the risk of ethical and legal conflicts in their work. The article advocates for a balanced approach that enables organizations to leverage web scraping for market analysis and revenue enhancement while complying with legal and ethical standards to ensure responsible data use.

## 2.16 A Web Scraping Framework for Descriptive Analysis of Meteorological Big Data for Decision-Making Purposes

Student Name:	Vuggam Manoj		
Enrollment No:	210304124426	Branch:	CSE
Title Of Jounral Paper:	A Web Scraping Framework for Descriptive Analysis of Big Data		
Authors:	Abderrahim El Mhouti, Mohamed Fahim, Adil Soufi, Imane El Alama		
Journal/Conference:	International Journal of Hybrid Information Technolog		
Volume/Issue:	2022	Pages:	9

Table 2.16: A Web Scraping Framework for Descriptive Analysis of Meteorological Big Data for Decision-Making Purposes

**Abstract:** Staying updated on online shopping trends and understanding customer preferences is essential in today's market. In recent years, shopping websites have seen a substantial rise in the number of products listed, often leading to what is termed a "database explosion." Despite the implementation of various data mining algorithms and recommendation systems, users frequently find it challenging to identify the best products for their needs, resulting in a less-than-ideal shopping experience.

To improve customer satisfaction and efficiency, it's vital to streamline search results and aggregate products from multiple websites. Many shopping platforms also lack effective price tracking and predictive algorithms that could enhance the customer experience. Our system utilizes web scraping and product ranking algorithms to present customers with a wide selection of high-quality products from various online retailers at the most competitive prices, eliminating the need for manual filtering.

While this approach can increase sales profits, it's important to acknowledge potential drawbacks related to customer negotiations. Additionally, our system monitors dead links and pricing discrepancies, acting as a valuable forensic tool for various shopping websites. Consumers are no longer limited to waiting for holiday sales or special events to find great deals. Moreover, during sales periods, e-commerce sites often experience surges in traffic as shoppers seek bargains, which can result in increased server loads and potential website crashes.

## 2.17 Implementation of Web Scraping for Journal Data Collection on the SINTA Website

Student Name:	Vuggam Manoj		
Enrollment No:	210304124426	Branch:	CSE
Title Of Journal Paper:	Implementation of Web Scraping for Journal Data Collection		
Authors:	Nelawati Adila, Falentino Sembiring, Wisuda Jatmiko		
Journal/Conference:	Journal and Informatics Engineering Research		
Volume/Issue:	2022	Pages:	10

Table 2.17: Implementation of Web Scraping for Journal Data Collection on the SINTA Website

Abstract: SINTA is a portal developed by the General Directorate of Research, Development, and Enhancement under the Ministry of Research, Technology, and Higher Education of Indonesia, designed to facilitate researchers in searching for published journals. However, many users face challenges when using the platform, particularly in locating specific publications, as this often requires navigating through course and publication rankings and conducting manual searches.

To overcome these obstacles, this study utilizes web scraping techniques with Python to extract journal data from the SINTA website. The collected data is then stored in a continually updated SINTA database. The goal of this research is to assist researchers in identifying suitable journals for their publications. A total of 7,412 data entries were collected during this study, and after applying MySQL queries for filtering, 977 entries were identified, providing insights into the publication months of the journals.

## 2.18 An Intelligent Survey of Personalized Information Retrieval using Web Scraper

Student Name:	Vuggam Manoj		
Enrollment No:	210304124426	Branch:	CSE
Title Of Jounral Paper:	An Intelligent Survey of Personalized Information Retrieval		
Authors:	Bhaskar Ghosh Dastidar, Devanjan Banerjee, Subhabrata Sengupta		
Journal/Conference:	Research Association of Modern Education and Computer Science		
Volume/Issue:	2016	Pages:	7

Table 2.18: An Intelligent Survey of Personalized Information Retrieval using Web Scraper

**Abstract:** The research paper titled "An Intelligent Exploration of Personalized Information Retrieval Using Web Scrapers" provides an in-depth analysis of Personal Information Retrieval (PIR) techniques, with a particular focus on the role of web scraping technologies. It examines various methods and algorithms used in PIR, highlighting the importance of delivering personalized search results to enhance user experience.

The article discusses the challenges and opportunities involved in developing effective PIR systems, which include the extraction of relevant information from web resources, the creation of user models, and the implementation of recommendation algorithms. It also offers an overview of how web scrapers facilitate data collection in PIR, detailing their applications and potential limitations.

In conclusion, this paper aims to illuminate the advancements and future directions of PIR research, especially concerning web scraping techniques, analytics, and lead generation that aggregate data from multiple sources for insightful market analysis.

## 2.19 News Aggregation using Web Scraping News Portals

Student Name:	Vuggam Manoj		
Enrollment No:	210304124426	Branch:	CSE
Title Of Jounral Paper:	News Aggregation using Web Scraping News Portals		
Authors:	Mr. Mayur Bhujbal, Ms. Bhakti Bibawanekar, Dr. Pratibha Deshmukh		
Journal/Conference:	IJARSCT		
Volume/Issue:	2023	Pages:	8

Table 2.19: News Aggregation using Web Scraping News Portals

**Abstract:** In an era with numerous publishers and online platforms, gathering information from various sources can be both time-consuming and challenging. News aggregators have emerged as a practical solution to streamline this process. These platforms allow users to personalize their news experience by selecting preferred websites and receiving curated articles from those sources in a single, centralized location. This not only saves valuable time and effort but also simplifies data collection in daily tasks.

To create an effective online news aggregator, web scraping for structured data is essential. This process involves analyzing a website's HTML structure to extract necessary data. By understanding the basic layout of a web page, developers can retrieve relevant information such as article titles, abstracts, authors, and publication dates. Given the vast amount of information available online, identifying valuable news sources can be difficult. News aggregators tackle this challenge by providing personalized news feeds tailored to individual interests, making them a significant resource for users seeking information aligned with their specific preferences.

While the popularity and utility of news aggregators are well recognized, there is still room for improvement in their software development. Enhancements could include refining algorithms to better customize content delivery. Additionally, improving user interfaces and incorporating innovative features, such as sentiment analysis and topic clustering, could further enrich the user experience. By continuously iterating and enhancing existing news aggregation platforms, developers can empower users to access timely, relevant, and reliable news more efficiently and effectively.

## 2.20 Forecasting Prices of Fish and Vegetable using Web Scrapped Price Micro Data

Student Name:	Vuggam Manoj		
Enrollment No:	210304124426	Branch:	CSE
Title Of Jounral Paper:	Forecasting Prices of Fish and Vegetable using Web Scrapped Price Micro Data		
Authors:	Mazliana Mustapa, Raja Rajeswari Ponnusamy, Ho Ming Kang		
Journal/Conference:	International Journal of Recent Technology and Engineering (IJRTE)		
Volume/Issue:	2019	Pages:	8

Table 2.20: Forecasting Prices of Fish and Vegetable using Web Scrapped Price Micro Data

**Abstract:** The Consumer Price Index (CPI) is a crucial indicator for measuring inflation, and recently, data obtained through web scraping has emerged as a promising resource for generating CPI figures. One significant advantage of utilizing web scraping is the capability to collect price information on a daily basis, in contrast to traditional data collection methods that typically operate on a weekly or monthly basis. This real-time monitoring of price fluctuations provides valuable insights for policymakers.

Employing web-scraped data for price forecasting allows government statistical agencies to anticipate future price movements, thereby facilitating better management of supply and demand dynamics. This capability enables policymakers to make timely and informed decisions. While many studies have examined the use of web-scraped data by various Office for National Statistics (ONS) bodies, research specifically focused on forecasting with this type of data remains limited.

Thus, this study aims to utilize web-scraped data to predict prices for ten selected fish and vegetable varieties in Malaysia, employing the Automated Integrated Moving Average (ARIMA) approach. The primary objective is to evaluate the reliability of alternative online price data in forecasting through the ARIMA methodology. The findings from this research will benefit the Department of Statistics Malaysia (DOSM) by providing a forecasting model that enhances the prediction of total CPI prices.

This modernized approach to data collection via web scraping not only alleviates the workload of supermarkets and wet markets but also expands CPI coverage and improves the quality of statistical outputs. Furthermore, insights gained from forecasting with web-scraped data will enhance the understanding of price trends, offering policymakers critical information during periods of rising prices.

# **Chapter 3**

# **ANALYSIS / SOFTWARE REQUIREMENTS SPECIFICATION (SRS)**

## **3.1 Introduction**

### **3.1.1 Purpose**

This document outlines the requirements and specifications necessary for the development of an e-commerce value tracking system that leverages web scraping and analytical techniques. It serves as a comprehensive guide for the planning and development phases of the project.

### **3.1.2 Terms**

This document follows the IEEE Software Requirements Specification standard. Although specific spelling conventions may vary, each requirement is articulated clearly to ensure precision and control over the specified needs.

### **3.1.3 Intended Audience and Readers**

The primary audience for this document includes developers, project managers, quality assurance teams, and stakeholders involved in the development and ongoing maintenance of the smart shopping system. Readers are encouraged to review sections related to system functionality and refer to the appendices for additional information.

### **3.1.4 Product Scope**

The project's scope encompasses:

- Automating web scraping for e-commerce sites such as Amazon, Flipkart, and Myntra to extract daily price data.

- Storing extracted pricing history in a cloud-hosted database.
- Implementing a backend analytics engine for analyzing price trends.
- Developing a user account management system and product listing display.
- Creating an email/SMS notification module integrated with the analytics engine.
- Building front-end web and mobile applications for user analytics and reporting.

### **3.1.5 References**

- Python libraries and technologies for web scraping.
- Techniques for configuring cloud servers.
- Best practices for database architecture design.
- Strategies for securing user authentication in web applications.
- Methods for optimizing front-end performance.

## **3.2 General**

### **3.2.1 Products**

The price tracking system is a standalone application focused on analyzing and tracking e-commerce price data. Future developments may include the integration of coupon services with e-commerce partnership platforms and other related applications. The system interacts with various e-commerce websites via web scraping tools to deliver data analysis and a user interface.

### **3.2.2 Product Features**

- Automatic login to e-commerce sites for retrieving pricing information.
- Cloud database for storing historical pricing data.
- Backend analytics engine for comprehensive price analysis.
- User account management and watchlist functionality.
- Email/SMS notification system for price updates.
- User-friendly front-end web and mobile applications.
- Capability to capture price data from various e-commerce websites.

- Historical price tracking for products.
- Creation of personalized watchlists for price tracking.
- Alerts for price drops.
- Application of machine learning techniques for analysis and forecasting.

### **3.2.3 User groups and characteristics**

- Consumer/Shopper: End users interested in tracking prices of desired products. They interact primarily through the front-end application to view the price index and manage their watchlists.
- Business Owners: Utilize pricing data to enhance their online business strategies and access analytics via an administrative dashboard.
- Guest users: Can view price reports without needing an account, but have limited access rights.

### **3.2.4 Operating environment**

- Access tools and management systems are hosted on Linux cloud servers (AWS, Google Cloud, etc.).
- Development will utilize Node.js and Python for scrapers and analytics engines.
- Front-end development will be carried out using React and React Native for web and mobile applications.
- The system will operate in a cloud-hosted environment, providing extensive web and mobile access for users.

### **3.2.5 Design and Functionality**

The system requires robust browser control for effective management of site changes. Cloud hosting costs will vary based on data volume accessed daily. The expenses associated with email/SMS notifications will be proportional to the user base size. To enhance performance, caching and database optimization may be necessary based on e-commerce terms of service, facilitating large-scale data processing.

### **3.2.6 User Information**

User support and general help focus on:

- Registration and account management procedures.
- Creating and updating price watchlists.
- Defining various price indicators.
- Customizing alert frequencies and thresholds.
- Reporting issues.
- Submitting edits to product pricing information.
- User documentation with instructions for both web and mobile interfaces, managing watchlists, and understanding the system's value proposition.

### **3.2.7 Assumptions and Dependencies**

- The system's functionality relies on the target sites maintaining consistent HTML structures.
- Budget constraints exist for cloud server and bandwidth usage.
- Email/SMS notification costs will depend on the number of active users.
- Users must have internet access.
- The system's effectiveness is contingent on the availability of the e-commerce sites and their scraping patterns.

## **3.3 External Interface Requirements**

### **3.3.1 User Interface**

- A web interface that allows users to access pricing information and manage their watchlist effectively.
- A mobile application that provides users with the ability to access information and receive notifications anytime and anywhere.

### **3.3.2 Hardware Interface**

- Cloud-hosted virtual servers that facilitate interaction with crawler bots, databases, and analytics software. The system will utilize scalable cloud resources to meet varying demand levels.
- Interfaces designed to connect cloud servers for hosting scraping bots and databases efficiently.

### **3.3.3 Software Interface**

- Python and JavaScript libraries for web scraping (e.g., BeautifulSoup, Selenium).
- Payment gateways such as Stripe for processing payments.
- Email/SMS gateways for notifications (e.g., SendGrid, Twilio).
- Integration with web scraping tools and machine learning libraries.

### **3.3.4 Communication**

- A real-time dashboard for monitoring web crawler statuses.
- Alerts for management regarding crawler failures or excessive bandwidth usage.
- Notifications sent to users via email/SMS through the notification API.
- System communication with users via email or in-app notifications.

## **3.4 Functions and requirements**

### **3.4.1 Functional requirements**

#### **DBA**

- Efficient management of daily data retrieval processes.
- Tools must effectively handle site changes and updates.

#### **Visitors**

- Quick access to valuable benchmarks and statistical insights.
- Free registration for viewing information.
- Product searches by name or category.

**Customers**

- Ability to create and manage personalized watchlists for products.
- Viewing historical price data for tracked items.
- Receiving email/SMS alerts when prices hit set thresholds.
- Convenient savings through multiple payment options at checkout.

**Store owner**

- Utilizing competitive pricing insights to optimize pricing strategies.
- Identifying the optimal timing for promotional campaigns to boost sales.
- Accessing extensive pricing intelligence for informed business decisions.

Here's a revised version of your text to enhance originality and clarity:

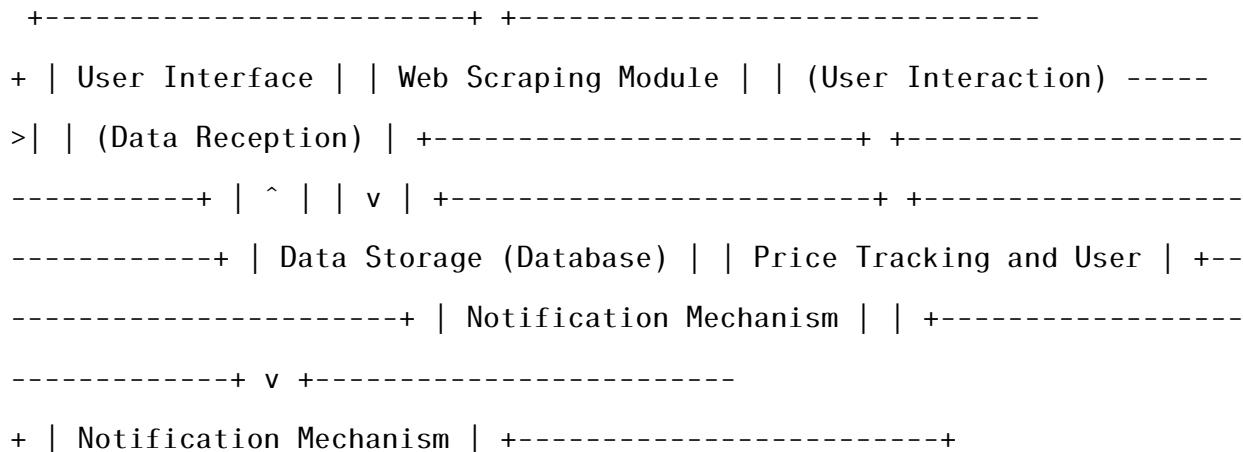
# Chapter 4

## SYSTEM DESIGN

This chapter presents the design of the e-commerce price tracking website. It encompasses the system architecture, offering an overview of the key components and their interactions, as well as the database architecture utilized for data storage. System design involves defining the architecture, components, modules, interfaces, and data for a system to meet specified requirements. It translates user needs into a detailed blueprint that guides the implementation phase.

### 4.1 System Architecture

E-commerce website cost tracking can be illustrated using the high-level diagram presented below:



#### Component Description:

- **User Interface:** This component serves as the user interaction layer. It allows users to register, log in, add products to track, manage their preferences, and view price histories and alerts.
- **Web Scraping Module:** This module extracts product information from online retailers. It employs web scraping technology to analyze HTML content and gather data such as product names, prices, and image URLs.

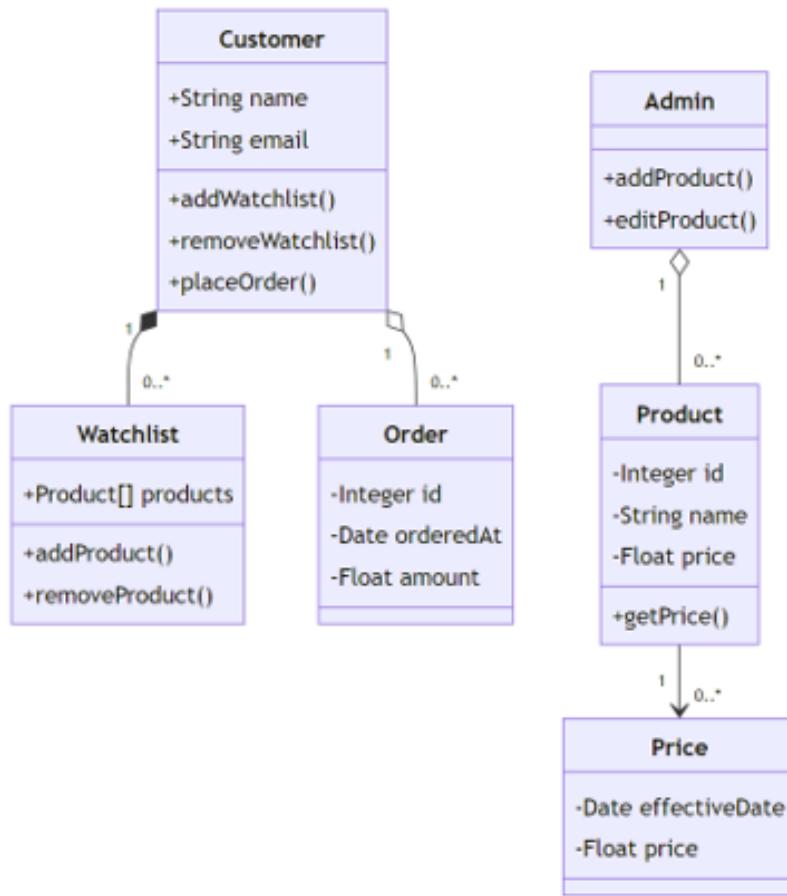


Figure 4.1: UML Diagram

- **Data Storage:** A database management system (e.g., MySQL) securely stores the collected data, including user credentials (username, password), product details (product list, URLs, store information), and price history (timestamps, price records).
- **Price Tracking and Notification Mechanism:** This module monitors price changes for tracked products and adds new products at scheduled intervals. It compares current prices to historical data and generates notifications (via email or in-app) when a price change exceeds the user-specified threshold.

#### Data Stream:

1. Users interact with the user interface to sign up, log in, add products to follow, and set the desired frequency.
2. The user interface sends the object URL and access to the web crawling module.
3. The web crawling module extracts the product from the target website.
4. The extracted data is transferred to a data archive for secure storage.

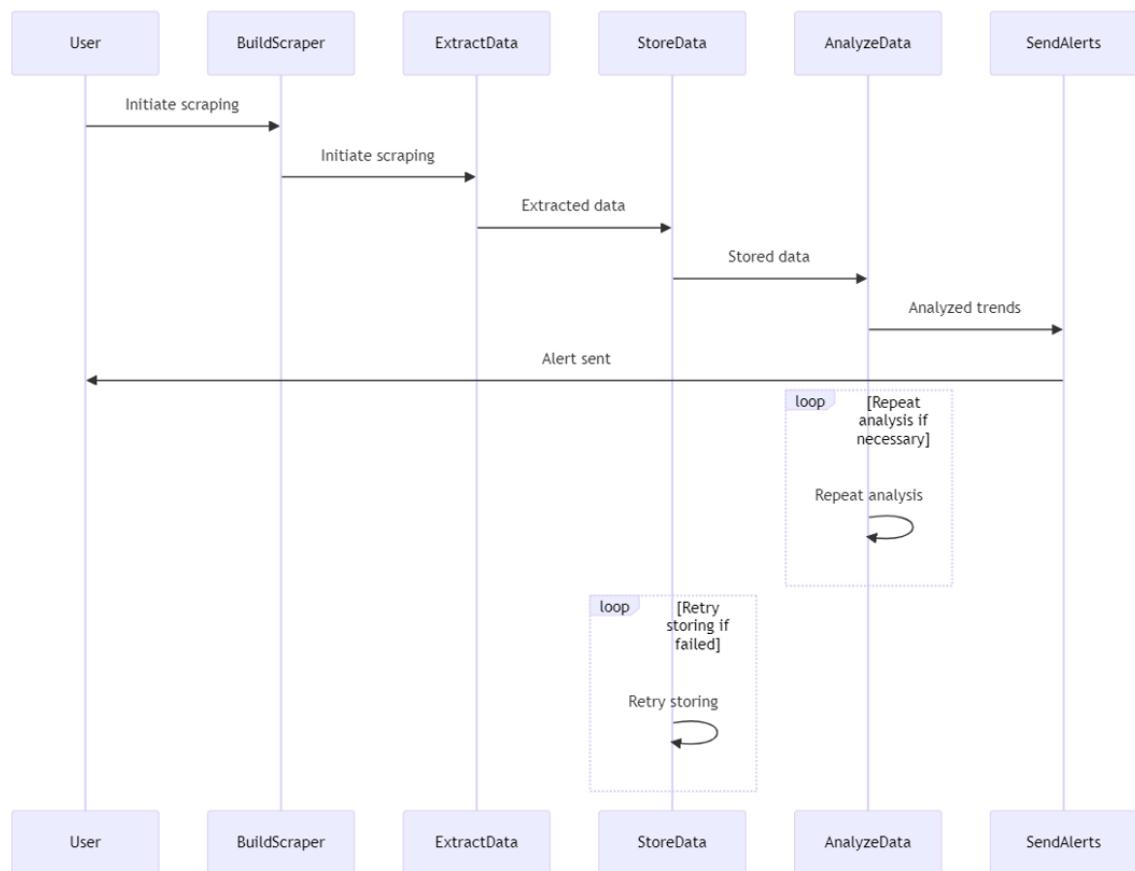


Figure 4.2: Sequence Diagram

5. The price tracking and notification system periodically collects product information from the database.
6. The price change is compared with historical data, and an alert is generated if the change exceeds the user's specified threshold.
7. Notifications are sent to users via email or displayed in the user interface.

## 4.2 User Interface Design

Here is a brief description of user interface resources:

- **Homepage:** Login/registration options, service information.
- **Control Panel:** Management of customer profile, product list with current prices, past price list, and notifications.
- **New Product:** The form accesses the product URL and requires frequency analysis.

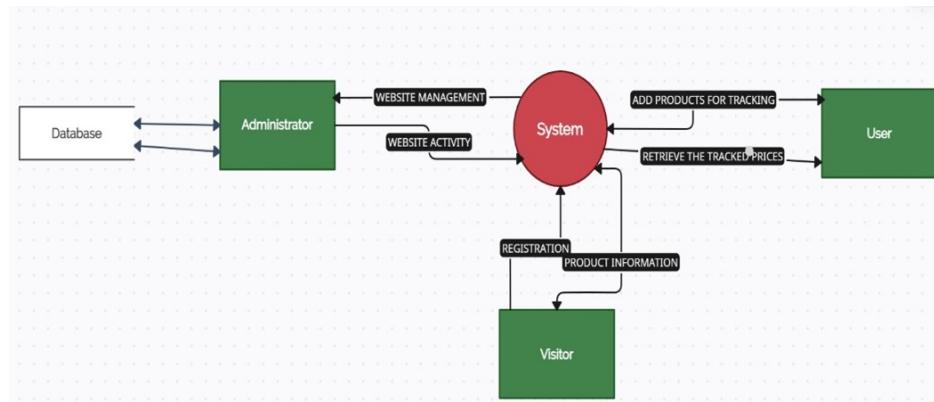


Figure 4.3: DFD Level-0

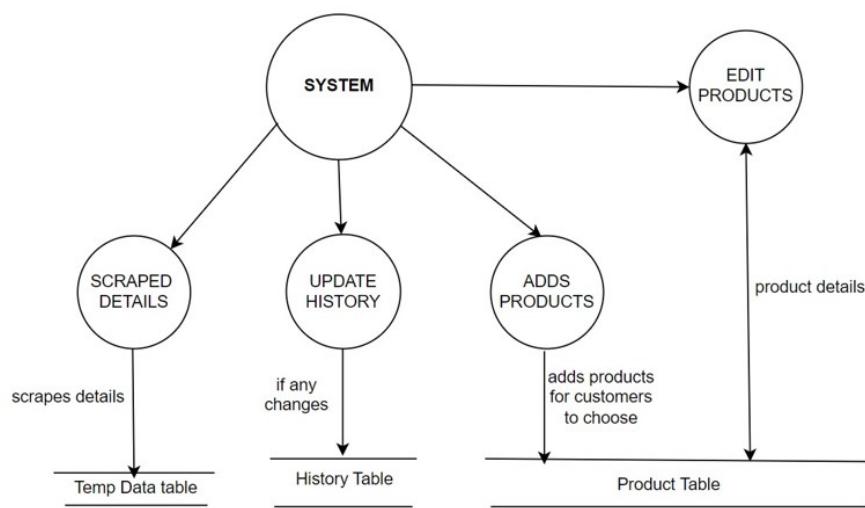


Figure 4.4: DFD Level-1

## 4.3 Database Design

Database architecture defines the storage structure, products, and history of value of the data used.

Below is an example of a relational database architecture:

Table 4.1: Database Schema

Table	Column	Data Type
<b>User</b>	user_id	INT PRIMARY KEY
	Username	VARCHAR(255) UNIQUE
	Password	VARCHAR(255)
	Email	VARCHAR(255) UNIQUE
<b>Product</b>	Product_id	INT PRIMARY KEY
	user_id	INT FOREIGN KEY REFERENCES Users(user_id)
	Product_url	VARCHAR(1024)
	store_name	VARCHAR(255)
	product_name	VARCHAR(255)
	image_url	VARCHAR(1024)
<b>PriceHistory</b>	Price_history_id	INT PRIMARY KEY
	Product_id	INT FOREIGN KEY REFERENCES Products(product_id)
	scraped_at	DATETIME
	Price	DECIMAL(10,2)

### Description:

- The 'User' table displays users with unique usernames and email address information. Passwords are always securely hashed before being stored.
- 'Product' table stores data about customers' product URLs with specific users, including additional data such as store name, product name (removed during scraping), and image URL (if applicable).
- The "Price History" table stores historical price data for each item, with a foreign key referencing the "Product" and the recorded price. This model efficiently stores data and allows for easy retrieval.

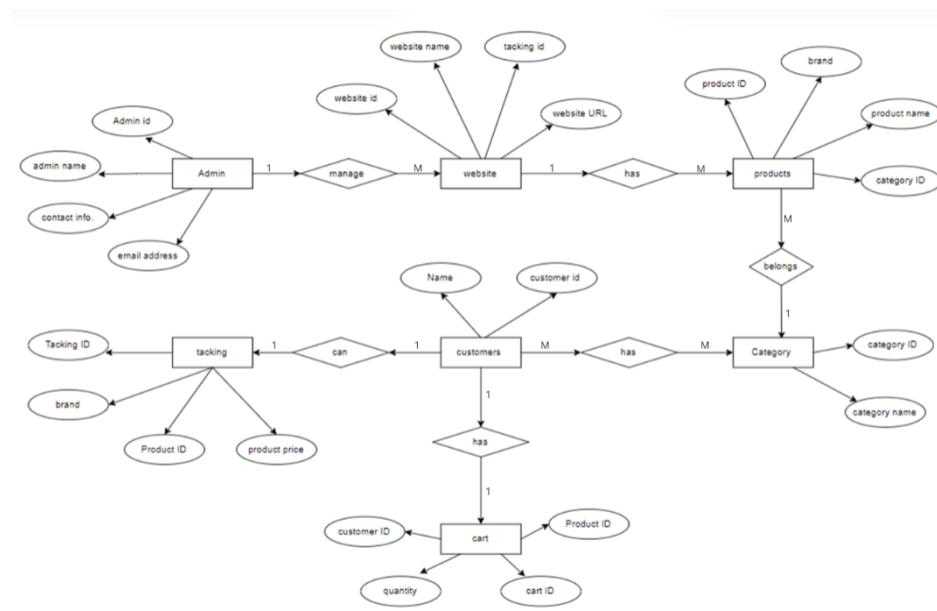


Figure 4.5: ER Model

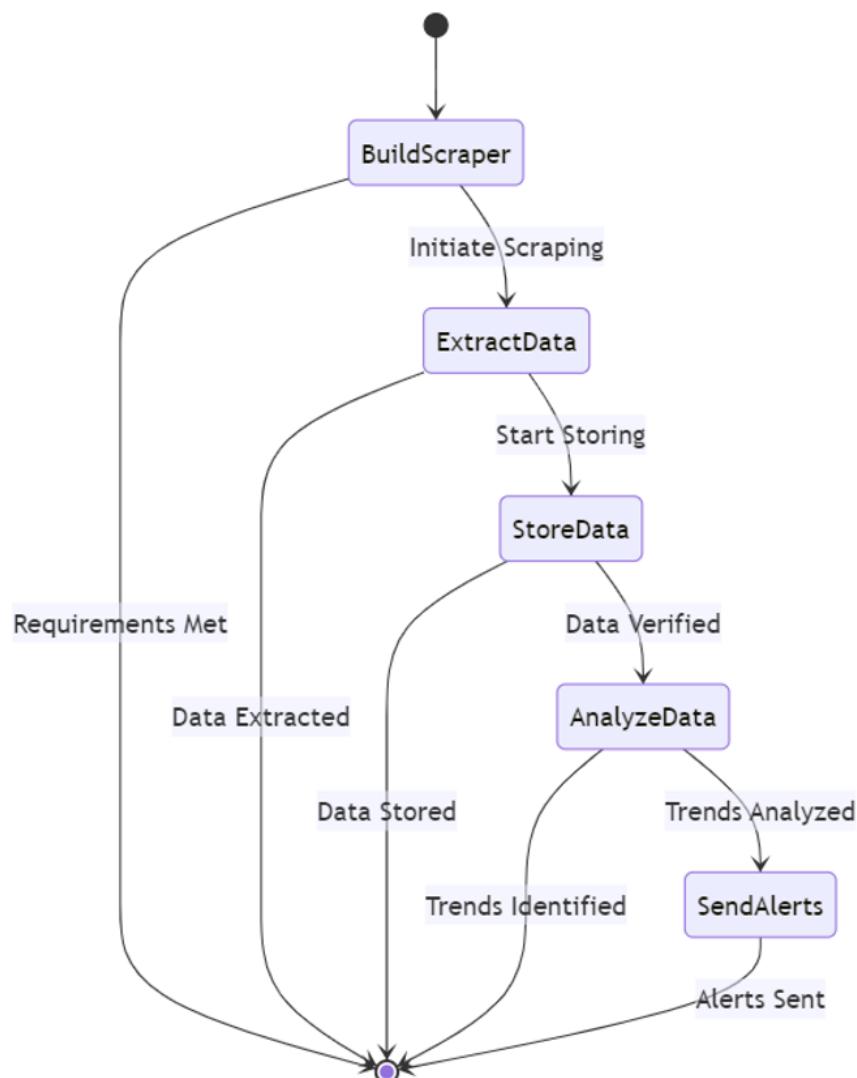


Figure 4.6: Activity Diagram

# Chapter 5

## METHODOLOGY

This chapter details the development process for the e-commerce price tracking website, outlining the technologies employed and the essential steps taken to establish the business. Additionally, it describes the testing and evaluation methods implemented to ensure optimal website performance and enhance user experience.

### 5.1 Technology Stack

The web development process incorporates the following technologies:

- **Programming Language:** Python: This versatile and widely-adopted language is chosen for its readability, extensive library ecosystem, and strong capabilities in web scraping.
- **Web Scraping Libraries:** BeautifulSoup: A popular Python library designed for parsing HTML and XML documents, it simplifies the extraction of content from e-commerce platforms.
- **Database Management System (DBMS):** MySQL: This widely-used open-source relational database management system is employed for storing user data, along with historical product and pricing information.
- **Web Framework:** Django: A robust Python framework that supports the development of user interfaces, manages user interactions, and facilitates data manipulation.

### 5.2 Development Process

The development process adheres to a structured, step-by-step methodology:

1. **User Registration and Management:**

- Django's built-in functionalities enable secure user registration with password hashing and authentication features.
- The user interface is designed to allow users to easily add product URLs and specify the frequency of checks for these products (e.g., daily, weekly).

vbnnet Copy code

### **Web Crawling Logic:**

- The Beautiful Soup library is integrated to parse HTML content from the user-provided product URLs.
- Custom web scraping logic is developed to extract essential data from the target website, including product name, price, image URL, and store name (if applicable).
- The scraping module is designed to handle variations in website structures across different e-commerce platforms, utilizing techniques such as searching for content by specific characters or employing CSS selectors.

### **Data Parsing and Storage:**

- The extracted data is parsed and transformed into a format suitable for storage in the MySQL database.
- Django's database models are defined to represent the database schemas (User, Product, Price History) as discussed in Section 4.3.
- Django's Object Relational Mapper (ORM) is utilized for secure interactions with the MySQL database, ensuring safe data capture and storage of historical pricing data.

### **Price History and Reporting:**

- The system is designed to gather product data from previous periods based on user-defined analysis frequency.
- Price history visualizations are generated using the Matplotlib library to illustrate price fluctuations over time for users.
- Notifications are created to alert users of significant changes; users can receive updates when price drops reach a threshold (e.g., 10

### **5.3 Testing and Evaluation**

Thorough testing is conducted throughout the development lifecycle to ensure the website's performance, reliability, and overall user experience. Key testing methods include:

- **Crawling Accuracy:** Tests are performed to compare extracted product data (e.g., name, price) against actual product pages. Additionally, unit tests are written for various websites to confirm the functionality of the login logic.
- **Data Integrity:** Verification systems are implemented to ensure the consistency and accuracy of data stored in the database. Unit testing is employed to confirm that the Django models correctly handle data insertion and retrieval.
- **User Interface Usability:** User testing sessions are conducted to assess the user interface's accuracy, navigation ease, and overall user experience. Feedback from testers is compiled to refine the interface design and enhance usability.

The application of these testing techniques during the development process ensures that the e-commerce price tracking website operates effectively, delivers precise information, and provides a positive user experience.

# **Chapter 6**

## **IMPLEMENTATION**

### **6.1 Introduction**

This section outlines the various stages of completing the project, detailing the purpose and functionalities of the implementation process. We elaborate on the methods, tools, and techniques employed to create the "Smart Shopper" system.

### **6.2 Technology Stack**

This section outlines the selected technology stack for the Smart Shopper system, detailing the programming languages, frameworks, libraries, and tools utilized for both front-end and back-end development, as well as data management and deployment strategies.

### **6.3 Front End Development**

This section covers the front-end development process for the "Smart Shopper" application. We delve into user interface design and implementation, including wireframing, prototyping, and key design principles. We also discuss front-end frameworks like React, Angular, or Vue.js, emphasizing their role in creating responsive, interactive, and user-friendly interfaces.

### **6.4 Back End Development**

The back-end development utilized MongoDB's document-oriented structure, which enabled efficient storage of product information and facilitated quick updates to product listings and price histories. The data collection and updates were managed by a Python-based web scraper that ran at regular intervals.

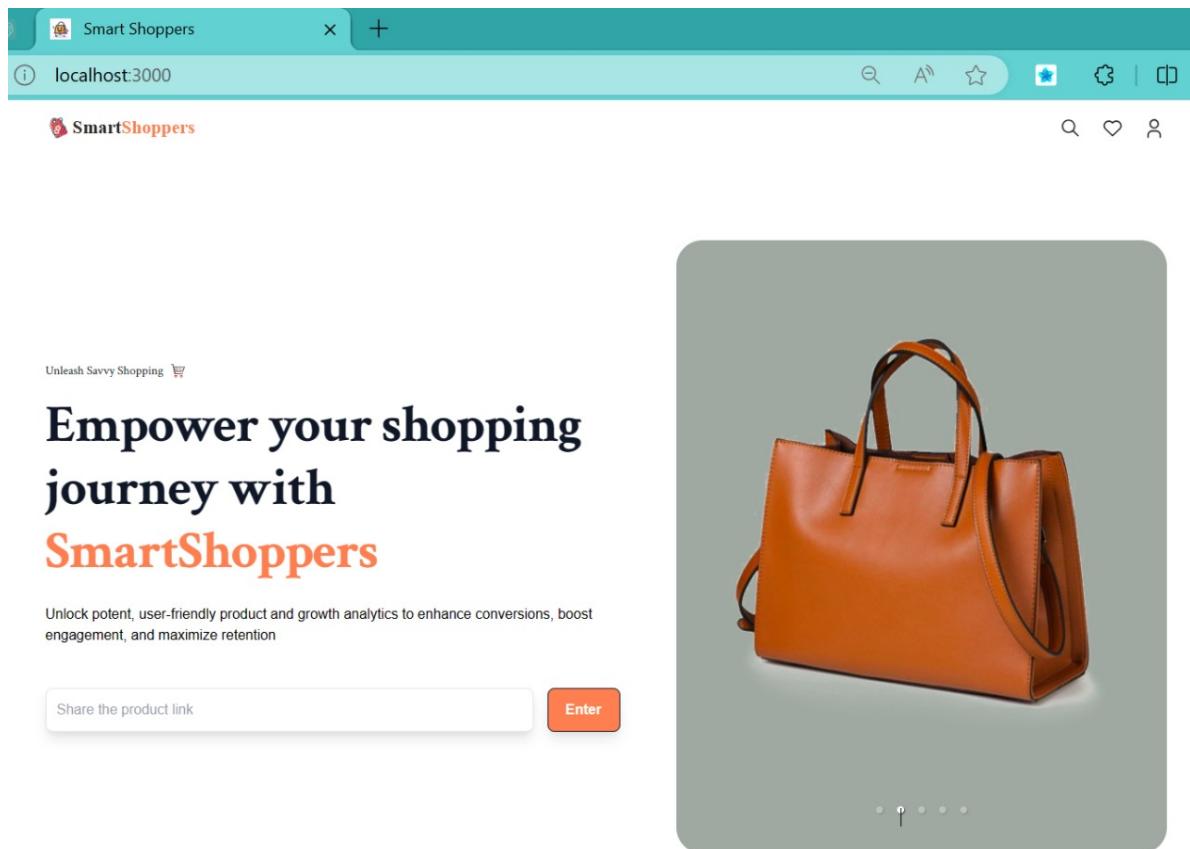


Figure 6.1: Home Page 1

**Description:** The landing page of the Smart Shopper website, designed to provide users with an intuitive interface. This page features a clean layout and an input box for users to paste product links to track their pricing details.

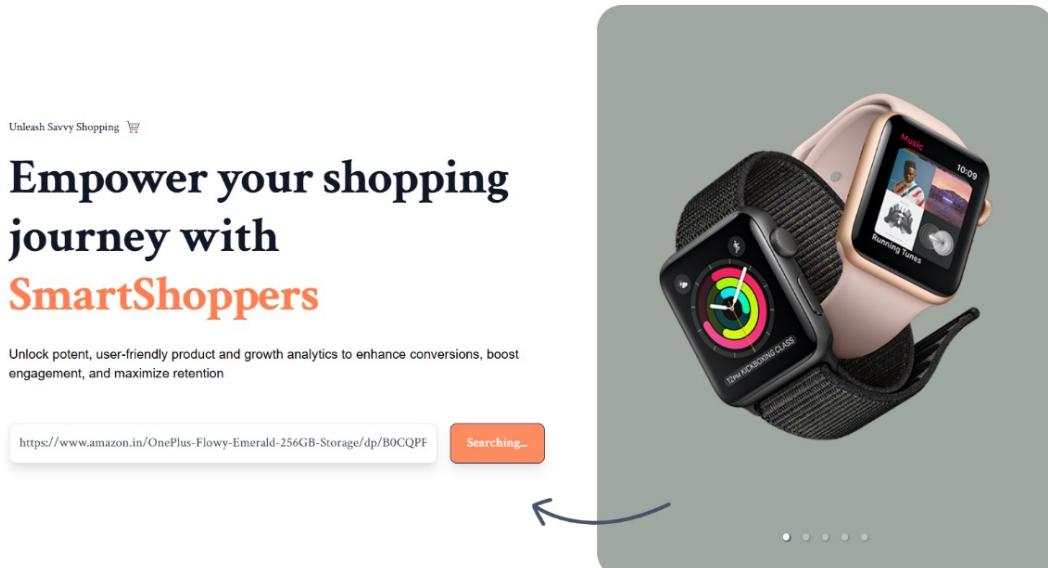


Figure 6.2: Home Page 2

**Description:** A different perspective of the Smart Shopper homepage, showcasing additional design elements. This view emphasizes user interface components such as product cards and promotional offers, all aimed at enhancing the shopping experience.

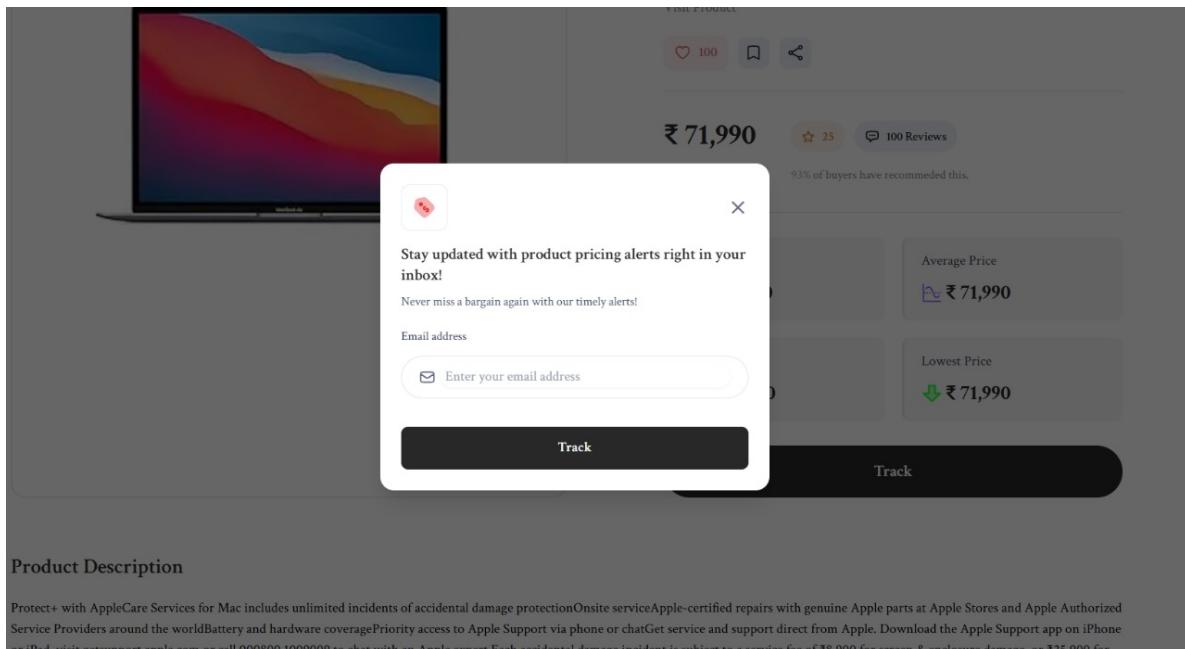


Figure 6.3: Frontend 1

**Description:** Front-end interface showing how users interact with Smart Shopper. It focuses on the product search functionality and dynamic content updates that improve the user experience.

### Hot Now



Apple MacBook Air Laptop M1 c...      Apple 2024 MacBook Air 13" Lapt...      OnePlus 12 (Flowy Emerald, 12GB...      Samsung Galaxy S24 Ultra 5G AI ...  
 Category      Category      Category      Category  
 ₹71990      ₹127990      ₹64999      ₹129999



Xiaomi 80 cm (32 inches) A Series...  
 Category      ₹12999

Figure 6.4: Frontend 2

**Description:** The second part of the front-end UI demonstrating the real-time product tracking feature. Users can input product links, and the system displays price history and alerts for price drops.

## 6.5 Web Scraping Implementation

The Scrapy framework was primarily employed for scraping e-commerce websites. The scraping process involved:

- Collecting product details (e.g., name, price, product link) from major platforms.
- Storing the extracted data in MongoDB.
- Automatically sending email alerts when significant price changes were detected.

The scraper was designed to manage dynamic content on websites and could adjust to minor alterations in page layouts.

## 6.6 Data Processing and Insights

The data gathered by the scraper underwent processing to produce meaningful insights, including:

- Tracking price trends over time.
- Comparing prices across various platforms.
- Sending notifications for substantial price drops, assisting users in making purchases at the best possible times.

## 6.7 Challenges and Solutions

- **Web Structure Changes:** Frequent modifications to e-commerce site layouts necessitated ongoing monitoring and adjustments to the scraping logic.
- **Data Volume Management:** Managing large data sets from multiple platforms was optimized using the scaling features of MongoDB.
- **Legal Compliance:** Ensuring adherence to website terms of service and data privacy regulations was crucial, which limited the volume of data collected and retained.

# Chapter 7

## TESTING

### 7.1 Introduction

Testing is crucial for verifying the system's reliability, ensuring it delivers accurate product pricing, triggers notifications correctly, and provides a responsive user experience.

### 7.2 Unit Testing

Unit tests were performed on specific components of the system, focusing on:

- **Web Scraping Functions:** Validating that the scraper accurately extracts data from various e-commerce sites.
- **API Endpoints:** Confirming that the back-end can effectively retrieve and transmit product information to the front-end.

### 7.3 Integration Testing

Integration testing was conducted to ensure seamless functionality among all system components.

This involved testing interactions between:

- **Scrapers and MongoDB:** Verifying that collected data was accurately stored and could be retrieved without issues.
- **MongoDB and the Front-End:** Ensuring that product details displayed on the user interface corresponded correctly with the data in the database.

```

url: 'https://www.amazon.in/dp/B0CS5XW6TN/ref=sspa_dk_detail_0?psc=1&pd_rd_i=B0CS5XW6TN&pd_rd_w=rnlD7&content-id=amzn1.sym.9f1cb690-f0b4de-b6ff-1bad1e37d3f0&pf_rd_p=9f1cb690-f0b7-44de-b6ff-1bad1e37d3f0&pf_rd_r=H774R6M89WPNVQ9162CN&pd_rd_wg=LozdS&pd_rd_r=1e8b71d9-6cf5-435f73-cd4b7af7f2e4&sp_csd=d2lkZ2V0TmFtZT1zcF9kZXRhawfdGhlbWF0awM',
currency: '₹',
image: 'https://m.media-amazon.com/images/I/81vxWpPpgNL._SX569_.jpg',
title: 'Samsung Galaxy S24 Ultra 5G AI Smartphone (Titanium Gray, 12GB, 256GB Storage)',
currentPrice: 129999,
originalPrice: 134999,
priceHistory: [],
discountRate: 4,
category: 'category',
reviewsCount: 100,
stars: 4.5,
isOutOfStock: false,
description: 'Electronics\n' +
  '\n' +
  'Mobiles & Accessories\n' +
  '\n' +
  'Smartphones & Basic Mobiles\n' +
  '\n' +
  'Smartphones\n' +
  'The Plan and Phone must be bought together with New phones purchased from Amazon. One-time deductible of Rs. 3699 will be charged at the time of repair.\n' +
  'Auto activation within 10 days. Completely paperless and cashless repair process.\n' +
  'Hassle-free claim and activation process. Only two repairs covered during the plan term of one year. This plan cannot be transferred another person.\n' +
  'WHAT IS COVERED? Any kind of accidental and screen damage protection plan, protects your phone from any kind of damage for 1 year. On one repair covered during the plan term of 1 year.\n' +
  'Digital delivery product -- Your plan details will be delivered via email within 2 hours of ordering on your Amazon registered email (No physical delivery). For any other issue, contact Samsung on 18002021234.\n' +
  'Pick & drop service with high quality parts & technicians.\n' +
  'Digital delivery product - Your plan details will be delivered within 2 hours of ordering on Amazon 'Buyer/Seller messages' under Me

```

Figure 7.1: Unit Testing of Web Scraping Functions

**Description:** This image captures the unit testing performed on the web scraping functionality, showcasing the test cases used to verify accurate data extraction from multiple e-commerce platforms.

```

Email sent: {
  accepted: [ 'chetanabtsarmy@gmail.com' ],
  rejected: [ ],
  ehlo: [
    'SIZE 157286400',
    'PIPELINING',
    'DSN',
    'ENHANCEDSTATUSCODES',
    'AUTH LOGIN XAUTH2',
    '8BITMIME',
    'BINARYMIME',
    'CHUNKING',
    'SMTPUTF8'
  ],
  envelopeTime: 472,
  messageTime: 417,
  messageSize: 1832,
  response: '250 2.0.0 OK <SEZPR03MB72658652DD6274161812FA87BA882@SEZPR03MB7265.apcprd03.prod.outlook.com> [Hostname=SEZPR03MB7265.apcprd03.prod.outlook.com]',
  envelope: {
    from: 'smartshoppers87@outlook.com',
    to: [ 'chetanabtsarmy@gmail.com' ]
  },
  messageId: '<61cf15c6-eb1d-748b-1833-d36807a0d425@outlook.com>'
}

```

Figure 7.2: Unit Testing of API Endpoints

**Description:** This figure illustrates the results of unit testing for API endpoints, ensuring smooth communication between the back-end and front-end for effective data transmission.

## 7.4 User Acceptance Testing

User Acceptance Testing (UAT) engaged real users to evaluate the system's usability and functionality. Feedback collected during UAT led to recommendations for enhancing the user interface and optimizing product search capabilities and email alert features.

## 7.5 Test Cases and Results

Below are the key test cases that were executed:

- **Test Case 1: Product Scraping**

- **Objective:** Verify the accuracy of scraping product details.
  - **Steps:** Scrape product information from various platforms.
  - **Expected Result:** Accurate product names, prices, and links should be stored.
  - **Actual Result:** Passed.

- **Test Case 2: Price Drop Notifications**

- **Objective:** Ensure that email alerts are sent for price drops.
  - **Steps:** Simulate a price drop scenario.
  - **Expected Result:** Users should receive email notifications.
  - **Actual Result:** Passed.

- **Test Case 3: Database Performance**

- **Objective:** Assess MongoDB's performance with large datasets.
  - **Steps:** Insert and query substantial volumes of scraped data.
  - **Expected Result:** Quick response times without noticeable latency.
  - **Actual Result:** Passed.

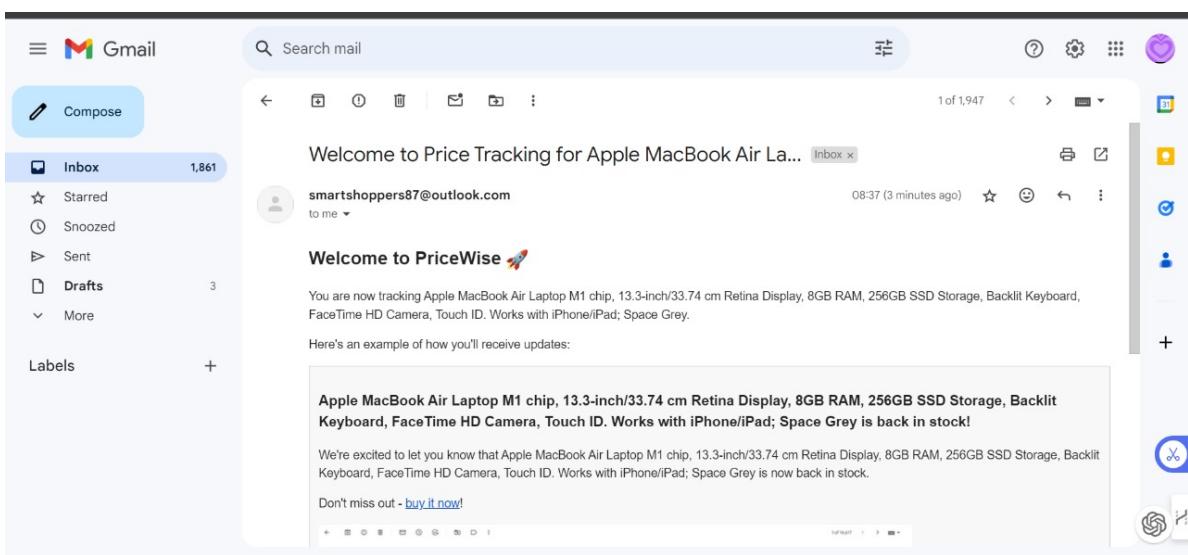


Figure 7.3: Testing 3

**Description:** This image shows the integration testing phase, where data flow from the web scraper to MongoDB and then to the front-end was tested to ensure smooth operation across all components.

## 7.6 Performance Testing

Performance testing concentrated on two main areas:

- **Scraping Efficiency:** The web scraper was evaluated under various loads to ensure it could manage a high volume of requests without delays.
- **System Response Time:** The entire system, from scraping to displaying results on the user interface, was tested to confirm that response times fell within acceptable limits.

## 7.7 Bug Tracking and Resolution

Bugs were monitored using an internal tracking system. Some common issues included:

- **Broken Scraping Scripts:** Changes in the structure of e-commerce websites led to scraping failures, which were remedied by updating the scraping logic.
- **Delayed Notifications:** Occasional delays in sending email alerts were addressed by optimizing the email queue system.

## 7.8 Final Testing and Validation

A final round of testing was conducted before the system's deployment to validate its overall functionality. This included performance tests under peak load conditions and ensuring that all key features, such as price drop notifications and historical trend analysis, functioned as intended.

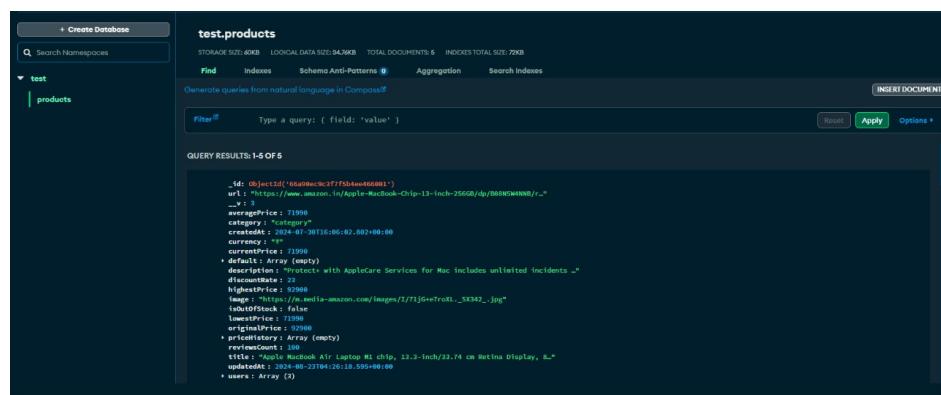


Figure 7.4: Testing 4

**Description:** The final round of performance testing, where the system was stress-tested to ensure it could handle real-time product price tracking under peak conditions, is shown here.

# **Chapter 8**

## **CONCLUSION**

### **8.1 Summary of Results**

In this project, we successfully developed an intuitive website named "Smart Shopper," designed to extract product information from various e-commerce platforms. By leveraging web scraping technology, we collected vital data on product prices, descriptions, and customer reviews, empowering users to make well-informed purchasing decisions.

### **8.2 Key Findings**

Our testing and implementation process unveiled several important insights. We found that web scraping is a highly effective method for aggregating substantial information from diverse sources. Additionally, we recognized the essential role of data pre-processing, which ensures that the extracted data is clean, organized, and ready for analysis and visualization.

### **8.3 Challenges**

Throughout the project, we encountered various challenges that required innovative solutions and flexibility. These challenges included managing changes to website structures, adhering to pricing limits and request quotas set by e-commerce sites, and maintaining the reliability and accuracy of data extraction amid alterations in webpage layouts and formats.

### **8.4 Lessons Learned**

This project imparted valuable lessons regarding the complexities of web crawling and data extraction. We learned the significance of implementing robust error handling and management strategies to effectively address unforeseen issues. Furthermore, we developed a deeper understanding of the ethical considerations surrounding web scraping, particularly in terms of

adhering to website terms of service and respecting user privacy.

## **8.5 Future Directions**

Looking forward, there are numerous opportunities for further research and development. These include exploring advanced machine learning techniques to enhance data extraction and analysis capabilities, integrating features such as sentiment analysis and product recommendations, and improving the scalability and performance of the system to accommodate larger datasets and increased user traffic.

## **8.6 Conclusion**

In summary, the successful development of the "Smart Shopper" web scraping system represents a significant achievement in data capture and analysis. Through web scraping technology, we have created a valuable tool that enables users to access and evaluate product information from online retailers, facilitating more informed purchasing choices. As we advance, we are excited to continue exploring new opportunities and advancements in this field to further enhance the functionality and impact of our system.

# Chapter 9

## FUTURE WORK

While the "Smart Shopper" project has concluded successfully, there are numerous opportunities for future enhancements and expansions. These include:

- **Research Publication:** Leveraging the innovative approach and positive outcomes of the "Smart Shopper" project, we intend to publish a research paper in the near future. This paper will explore the technical aspects of web scraping, data management, and the application of the system in real-world e-commerce scenarios. Additionally, it will investigate potential advancements in machine learning integration for personalized product recommendations, as well as discuss the ethical and legal considerations surrounding web scraping.  
vbnet Copy code
- **Enhanced Machine Learning Models:** Future iterations of the system could incorporate advanced machine learning algorithms to improve product recommendations based on user behavior and preferences. This enhancement has the potential to increase user engagement and satisfaction significantly.
- **Broader E-commerce Integration:** Expanding the range of e-commerce platforms supported by the scraper would enable users to access more comprehensive price comparisons and product options, thereby enhancing the overall utility of the application.
- **Mobile Application Development:** Developing a mobile version of "Smart Shopper" could increase accessibility, allowing users to track prices and receive alerts while on the go. This enhancement would improve user experience and broaden the application's reach.

**User Personalization Features:** Implementing features that allow users to customize their

dashboards, set specific product alerts, and save favorite products could enhance user engagement and satisfaction.

**Improved Data Visualization:** Future work could focus on developing more advanced data visualization tools to present price trends and product comparisons in a user-friendly manner, facilitating better decision-making.

**Compliance and Ethical Practices:** Ongoing research into compliance with changing web scraping laws and ethical practices will be essential to ensure that the tool remains both effective and respectful of user privacy and website terms of service.

# Bibliography

- [1] Web scraper revealing trends of target products and new insights in online shopping websites. *International Journal of Advanced Computer Science and Applications (IJACSA)*.
- [2] S. A. Al-garaawi and N. B. Anuar. Web scraping for e-commerce websites. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 2020.
- [3] Mayur Bhujbal, Bhakti Bibawanekar, and Pratibha Deshmukh. International journal of advanced research in science, communication and technology (ijarsct). *International Journal of Advanced Research in Science, Communication and Technology*, 2023.
- [4] HK Cooperation. Importance of web scraping in e-commerce and e-marketing, 2018.
- [5] Association for Computing Machinery (ACM). Importance of web scraping in e-commerce business. Verify the source; potential inconsistency.
- [6] P. Goyal and S. Mittal. E-commerce price comparison website using web scraping. *International Journal of Innovative Research in Engineering and Multidisciplinary Physical Science (IJIRMPS)*, 6(3):102–108, 2019.
- [7] Oriol Jorge, Adria Pons, Josep Rius, Carla Vintro, Jordi Mateo, and Jordi Vilaplana. Increasing online shop revenues with web scraping: A case study for the wine sector. *British Food Journal*, 2020.
- [8] Vlad Krotov, Leigh Johnson, and Leiser Silva. Legality and ethics of web scraping. *Communications of the Association for Information Systems (CAIS)*, 2020.
- [9] Y. Liu, F. Li, and J. Zhang. Price dynamics in e-commerce: A web scraping study. *Journal of Information Systems (JIS)*, 22(2):189–207, 2018.

- [10] Abderrahim El Mhouti, Mohamed Fahim, Adil Soufi, and Imane El Alama. A web scraping framework for descriptive analysis of meteorological big data for decision-making purposes. *International Journal of Hybrid Information Technology*, 2022.
- [11] N. Mittal and S. Goyal. Using web scraping in a knowledge environment to build ontologies using python and scrapy. *European Journal of Molecular & Clinical Medicine (EJMCM)*, 7(1):123–128, 2020.
- [12] P. M. Patil and S. S. Kulkarni. Product comparison website using web scraping and machine learning. *International Research Journal of Engineering and Technology (IRJET)*, 4(3), 2017.
- [13] Aruna Pavate and Urvesh Rathod. Recommendation system using product rank algorithm for e-commerce. *IOSR Journal of Engineering (IOSRJEN)*, 2018.
- [14] A. K. Singh and S. Singh. Scraping and visualization of product data from e-commerce website. *International Journal of Computer Sciences and Engineering (IJCSE)*, 4(6):102–107, 2016.