Product Comparison and Sentiment Analysis using Web-Scraping Techniques

Saurav Singh

Artificial Intelligence and Data Science
Thakur College of Engineering and
Technology,
Mumbai, India
Saurav2021c@gmail.com

Vishal Gupta

Artificial Intelligence and Data Science
Thakur College of Engineering and
Technology,
Mumbai, India
Vishalrg02@gmail.com

Mithil Kadam

Artificial Intelligence and Data Science
Thakur College of Engineering and
Technology,
Mumbai, India
Mithilskadam02@gmail.com

Abstract- Cost comparison site have ended up nowadays one of the foremost vital sources for the buy of all sorts of items. Numerous techniques were created by analyzing the user's behavior in arrange to pull in more commerce and lock in individuals and using opinion investigation we are able more in-depth information approximately clients intrigued. Since there are a lot of e-commerce sites available, it gets to be troublesome for clients to select the leading cost for the specified item among those destinations. Comparing e-commerce items utilizing web pilgrim permits clients to analyze prices and get the specified item at the least price. Users can moreover select a few items inside the same category to compare their characteristics. To induce data around items from e-commerce web crawlers and web scraping techniques are utilized to induce nitty gritty data. The proposed framework is based on users' desires and security. The client can clearly know the precise cost of the item and is trying to find offers completely different ecommerce destinations. The proposed framework can be a straightforward and viable way to know the precise cost and the offers given by the diverse e-commerce websites. This way points to supply a arrangement for online clients to buy items at cheap costs and spare their profitable time, exertion and money

File Terms- web mining estimation, examination, web scratching, cost comparison, e-commerce

I. INTRODUCTION

A cost comparison site acts as a stage or medium between shoppers and venders. It permits the buyers to see distinctive cost records for the item chosen by the client and helps the buyers to form an educated choice on which one to select in arrange to spare cash. It also serves as a apparatus to assist shoppers increment their cost mindfulness so they do not feel deceived by notices from retailers who claim they offer the most reduced cost, but the reality is different.

According to Social, Computerized and Portable in India investigate by We Are Social, web infiltration in India is 59% and the normal number of hours Indian web clients spend utilizing the web each week is 19.8 hours. In the interim, 21% of Indian internet clients get to the web by means of versatile gadgets, which implies that they can get to the internet anyplace with their smartphones. 77% of Indian web clients shared their considerations almost a brand through social media during this think about conducted in 2019. The study shows how associated people in India are to the Internet.

As mentioned above, a price comparison website acts as a medium between customers and retailers to allow customers to shop online at specific retailers to shop that offer such services. In other words, the price comparison website also has the function of promoting the retailer/store/hypermarket/supermarket to customers. Time and money pressures, especially in today's economy where the cost of living is rising and there's time for home shopping, a price comparison site like PriceWar.com is sure to be of great help to consumers. In addition, today's users are very comfortable with the Internet having developed a greater variety of applications. of networking and now provide users with various references.

II. LITERATURE REVIEW

The Primary inquire about analyzes the importance of cost comparison locales in upgrading advertise effectiveness and cultivating cost competition. These stages serve as important aggregators of item data, rearranging the consumer's look handle. Within the past, customers depended on manual strategies like books and catalogs, which were time-consuming. Cost comparison destinations, presently prepared with progressed innovation, offer highlights such as fetched calculation, web slithering, and moved forward client control. They collect and show costs, items, and important information from third parties, streamlining the decision-making prepare for buyers [1].

The Moment Investigate consider on UK Common Protections sector's cost comparison websites appears expanded promoting spend and competition, contrarily affecting money related execution. Be that as it may, tall client numbers and location utilization demonstrate a maintained future for these websites. To compete online, major brands require noteworthy showcasing budgets, making it challenging for littler players. A few comparison locales, like "Compare the Advertise," utilize

imaginative campaigns to pull in clients and boost their brand affect and site activity [2].

The point of the third paper is to created the 'modified SentiWordNet algorithm' that mainly is on machine learning is utilized to encourage the choices of the stock. These highlights will at that point be keep and may be created out there for the clients though they're searching for the stock. This approach offers the rating at the tip for each feature of the stock and so recommends the clients to travel with the one that has the finest evaluated item in an awfully express highlight. At whatever point the need of the choices is chosen, then the items that have most noteworthy positive scores in that highlights would be counseled for the client and hence makes a difference in shopping for the stock in line with their prerequisites. The framework makes a difference the makers of the stock to get it their criticism of the stock and so makes a difference them make strides the genuine alternatives of the stock and create stock that are steady with the necessities of the client [3].

This Fourth paper contributes a very particular approach for the robotized recognizable proof and extraction of item worth information from rash e-shop websites that's independent from the e-shops' dialect and after that stock space. The development of the approach compared to blessing approaches and devices for net information extraction and cost comparison is that the freedom from the dialect and so the item space of the e-shops, the direct setup of the extraction errands in expansion as a result of the control to naturally recognize the stock records among a total e-shop location. The approach employments tag way examination. The approach was made independent from a item domain or a dialect [4].

In this Fifth paper, the most quality characteristics of ecommerce site were portrayed. The e-commerce industry is quickly developing all showcasing communication blend of e-commerce organizations. It complements coordinate offering exercises, presents supplemental fabric to customers, projects an organization picture, and provides basic data to clients. Hence, it is critical for e-commerce organizations to get it client prerequisites, and at that point to make strides the quality of site appropriately. The most objective is to establish quality requirements to arrange the list of characteristics that may well of a quantitative evaluation, comparison and ranking process. Future work includes the benchmarking process against exemplary or well-known sites, such as Amazon.

III. METHODOLOGY

1.1 Proposed Work

Stage 1: Project Planning

In this initial phase, the research problem is identified, and the study's significance is established. The project's objectives and scope are defined, ensuring the feasibility of the work within the given timeframe. The methodology for addressing the problem is examined, and the system types to be developed and tools for system creation are identified through literature reviews.

Stage 2: Data Gathering and Analysis

Extensive research is conducted to further understand energy consumption and power. Readings are taken to gain insight into the concept of comparison websites, how they aid in problem-solving before purchasing home grocery products, and examples of existing comparison sites to assess the competition. Information from previous studies is obtained from scholarly research papers and is detailed in the literature review section. User data is collected through online surveys and face-to-face interviews with respondents, including store owners in the Perak area.

Stage 3: Research on any Existing Similar Systems

The next step involves investigating the existence of similar systems. The primary purpose of studying similar existing systems is to understand their functioning, the underlying concepts, calculations performed, and how they contribute to problem-solving.

Stage 4: Drafting the Main Components of the System

Following the research on comparable systems, the main components of the system to be developed are identified. Users must register with basic information such as name and email to access the website's services. Registered users are automatically subscribed to the site's newsletter. Users can select products, and related information will be displayed. Users can also add preferred items to their profiles, and current prices for those items will be sent separately, in addition to daily promotions. The main components of this system include a database for storing products and user information, and users can search for the products they are interested in.

Stage 5: Develop System Architecture

In this phase, the system's architecture is developed to provide a clear understanding of how the system will operate and to prevent the creation of a system that does not address the intended problem.

Web Crawler: The system involves a price comparison engine, and to collect a substantial amount of data from various e-commerce websites, an automated web crawler is necessary. Manually collecting data from websites is impractical, so an automated web crawler is used to navigate these e-commerce sites. The retrieved URLs are sent to the scraper for data extraction.

Web Scraper: Web scraping is used to extract HTML data from URLs and utilize it for personal purposes. Given that this is a price comparison website, data is scraped from multiple e-commerce sites. In this system, scraping is performed using Python libraries like "requests" and "beautifulsoup4." Beautifulsoup4 is used for parsing HTML pages. Using these tools, product information from different e-commerce websites is scraped and stored in the database.

Flask is a Python Web Framework: The product and price comparison website, created using the Flask framework, allows users to query products through SQL database using an object-relational mapper (ORM). The steps to submit a research paper to a journal are listed below.

Figure: System Architecture

Figure: Web Scraping Storage Fow

This Diagram Explains how we can pass the query and by which process and I which formats the data will be stored after research review of existing systems.
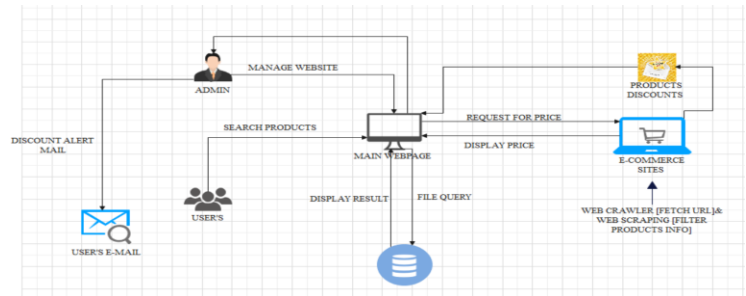

Figure: Workflow of Comparisons

The backend framework comprises of two imperative procedures web slithering and web scratching. Web scratching may be a that's utilized to extricate data in a human-readable organize and show it on the goal terminal. But some time recently rejecting the yield, Web Crawlers are responsible to explore to the goal once the crawler comes to the right page and matches up with the items, the scrapping process begins. Web rejecting basically comprises of two assignments: the primary is to stack the specified web page and the second is to parse the HTML data of the page to find the expecting data. In this framework, scratching is done utilizing Java because it gives a wealthy set of libraries to address these errands. "requests" is utilized to stack the URLs, and the "jsoup" library is utilized to parse the internet page as After rejecting the item data from different e-commerce websites, the information is shown on the site frontend comprising of the Most site. The clients explore for the desired item within the look bar and the inquiry is terminated within the nearby database i.e., sqlite5. The site is outlined utilizing jsp and jsoup which are composed in Java. Required comes about are recovered and shown on the Most site. The client can at that point compare the costs of items that are accessible on e-commerce websites. As before long as the client chooses a moo cost, he will be diverted to the initial e-commerce site. The items & user's information are secured by AES calculation as appeared in figure .Another include given could be a alert, in which the user can look for a item in n number of times on the off chance that the item has any rebate on it. at that point the admin can send the markdown take note message to the email-id user.

1.2 Techniques Used

Hardware Requirements:

System   : Pentium IV 2.4 GHz.
Hard Disk   : 40 GB.
Monitor   : 15 VGA Color.
RAM   : 2 GB.

Software Requirements:

Operating System : Windows XP/7.
Coding Language : Python, Flask.
Front End   : HTML, CSS, Tkinter, Flask
Database   : SQLite3.
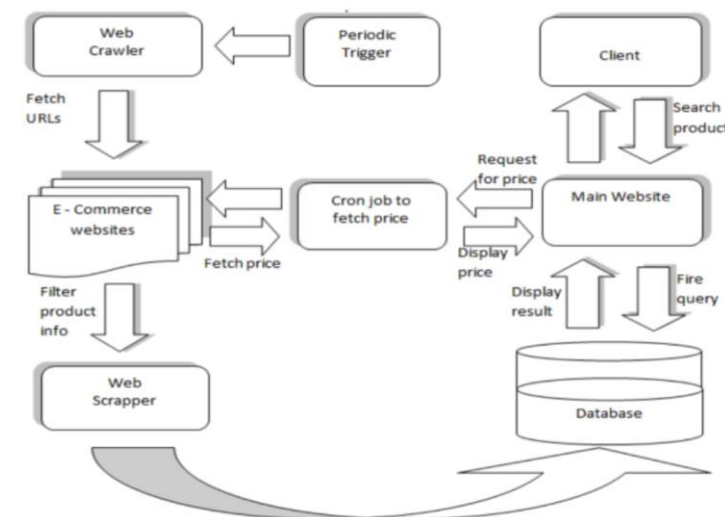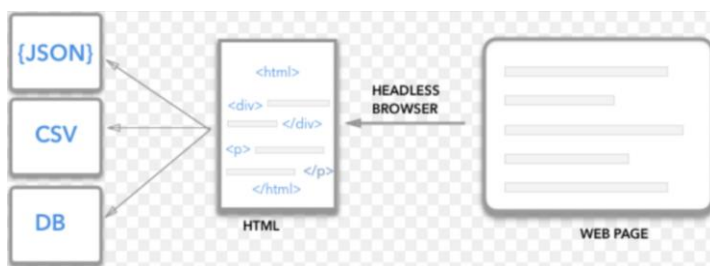
IV MODULES

1. Login Module

This module enables users to log in to the system and search for products.

2. Admin Panel

In this module, users have admin privileges, allowing them to add new product details and the corresponding URLs of e-commerce sites.

1. Live Product Price Comparison

Users can perform product comparisons across various sites such as Amazon, Flipkart, Snapdeal, and more in this module.

2. Sentiment Analysis on Reviews Using Machine Learning

This module integrates machine learning algorithms to assess the sentiment of product reviews.

3. Sellers/Retailers

Parties interested in partnering with PriceWar.com can use this module to submit price lists to the site's admin. They can also notify the admin about new products, sales, and promotions in their stores.

IV. CONCLUSION

A The proposed enhancements to PriceWar.com introduce a valuable feature for users: the ability to set price trigger alerts. Users can specify a target price for a particular product, and when a seller matches this price, the system sends an email notification, similar to alerts in stock exchange tools. This web-based system is primarily focused on e-commerce product comparison and aims to empower users with informed online shopping decisions. It allows users to analyze prices across various e-commerce platforms, ensuring they secure the best deals.

Our platform facilitates product comparison based on specifications within the same category, saving users both time and effort. It aggregates the strategies and top offers from major e-commerce stores, simplifying the decision-making process and enabling shoppers to make well-informed purchases online. This addition enhances the platform's utility, making it a valuable tool for cost-conscious online shoppers. Price Comparison using Sentiment Analysis makes it more reliable source of making comparisons of good products at a time. And it will provide overall optimized comparisons of product in real-time.

REFERENCES

[1] Pang, B., Lee, L. and Vaidyanathan, "S.Thumbs up? sentiment classification using machine learning techniques," Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10 (pp. 79-86). Association for Computational Linguistics, July 2002

[2] Webcredible Knight, E. (2017) : "The Use of Price Comparison Sites in the UK General Insurance Market.Moraga-Gonzalez, J.L.

[3] Priyanka M B, Subham Ankit Karan, Pruthvi M, Manushree M R and Sharath Kumar S," Recommender System Using Machine Learning" 2019

[4] Dao-Quan Li, Hua Yang and Li-li Li," Research and Application of Data Mining Technique in E-commerce" 2010 International Conference on E-Business and E-Government.

[5] Ashraful Alam, Atqiya Abida Anjum, Fahmid Shafat Tasin, Mizanur Rahman Reyad, Sadia Afrin Sinthee and Nahid Hossain," Upoma: A Dynamic Online Price Comparison Tool for Bangladeshi E-commerce Websites" 2020 IEEE Region 10 Symposium (TENSYMP)

[6] Nitha, Velayudhan,Abinav Vijay P,Alisha P D,Chithira Remesh, Lakshmi C V:E COMMERECE WITH PRICE COMPARISON,PRICE ALERT AND FAKE REVIEWDETECTION, In 2021 International journal of creative research thoughts.

[7] FAN LIANG, WEI YU, DOU AN, QINGYU YANG, XINWEN FU,AND WEI ZHAO:A Survey on Big Data Market: PricingTrading and Protection ,In 2018 IEEE Conference paper.

[8] Ahmad pourmini ,shashramnasiri : Web content extraction using contextual rules, IN 2015 IEEE International Conference on knowledge based Engineering and innovation(KBEI)

[9] Aravind Kumar M, Muthu Selvam M, Logeshkanna V: Online Shopping Comparison on E-Commerce Sites Using Web Scrapping Approach, In 2020 Journal paper on data mining.

[10] Mohammad Daoud,S.K Naqvi,Asad Ahmad :Opinion Observer: Recommendation System on ECommerce Website, In 2014 , International Journal of Computer Applications.

AUTHORS

First Author – Saurav Singh, Student in dept. of Artificial Intelligence & Data Science

Second Author – Mithil Kadam, Student in dept. of Artificial Intelligence & Data Science

Third Author – Vishal Gupta, Student in dept. of Artificial Intelligence & Data Science