International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

Monitoring Best Product using Data Mining Technique

Mamta Yadav[1], Ruturaj Chaudhari[2], Shreyas Pandey [3], Prof. Dilip Dalgade[4]

[1-4] Department of Computer Engineering, Rajiv Gandhi Institute of Technology, Mumbai, Maharashtra, India

---

Abstract - *E-commerce is widely increasing business in the world with increasing revenues every year by manifold times. This is simple indication that more people each day are moving on-line for shopping. It suggests that the preference of the people is changing. People prefer to buy products more and more from online e-commerce websites. They are successfully able to bring the customer in confidence and overcoming the risk involved in online transactions by carefully analyzing the behaviour of the buyers/sellers and cultivating confidence among resellers and buyers by developing strategies to attract more business and participation from people. Here comes the existence of price comparison engines into picture which enables the buyers to compare products from different e-commerce websites thus facilitating the buyers to purchase the product at the cheapest price. To obtain prices and other parameters of the products from different e-commerce websites, a web Crawler is used that can crawl to different e-commerce websites. To and fetch the URLs of the products. Once the URL of the products is available scrapper scrapes the details that were abstracted within the URL. This way the Project aims to provide a solution which grants power in the hands of the users to purchase genuine products at genuine prices and saving user's time, money and efforts.*

Key Words: Web Crawler, Web Scraper, Beautiful Soup, MongoDB, Elastic Search.

1. INTRODUCTION

As we are in the world of technology now a days consumers have an option of shopping or we can say buying various stuffs online from an E-commerce websites over the internet. But the main focus is to get the desired product in minimum price from the various E-commerce websites they survey, but all those e-commerce websites do not satisfy the consumer's needs due to limitations they impose. Monitoring best product using data mining technique overcomes this issues, using this technology the consumer can get the desired product by comparing them at one place and then finalizing the product based on the price which fits the budget. In Monitoring best product using data mining technique, intelligent agent is used which crawls through various websites and fetches URLs of different products, this intelligent agent is called as Web Crawler which is an automated program. Most e-commerce websites show products which are available with them at a particular rate. Using this technology consumer will have an upper hand in choosing the product as the details are available at one place which will also save time, money and efforts of the user and because of this the fact of visiting the store physically gets eliminated.

1.1 WEB CRAWLER

Web crawler is one of the main component of the Project. Since the product is price comparison engine, the first thing that is required is to collect large amount of data in terms of products from different e-commerce websites. Manually, the collection of such large amount of data was not possible. So the best way to get these data is to create a web crawler also known as spider. The main purpose of the crawler is to crawl different e-commerce websites and to fetch the URLs of the products from these websites. Then the fetched URLs that we have are given to the scraper for scraping purpose.

1.2 WEB SCRAPER

Web scraping can be defined as a process of extracting HTML data from the URLs and then using this data for personal purposes. Once we have the fetched URLs with us then the job is to get the information that is abstracted within the URL. These information are to be extracted for the purpose of comparison. The scrapper scrapes the information on this page on the basis of the tags in which the element. In this way the information can be extracted that are abstracted within the URLs. The extracted information is then stored in the database in the unstructured format.

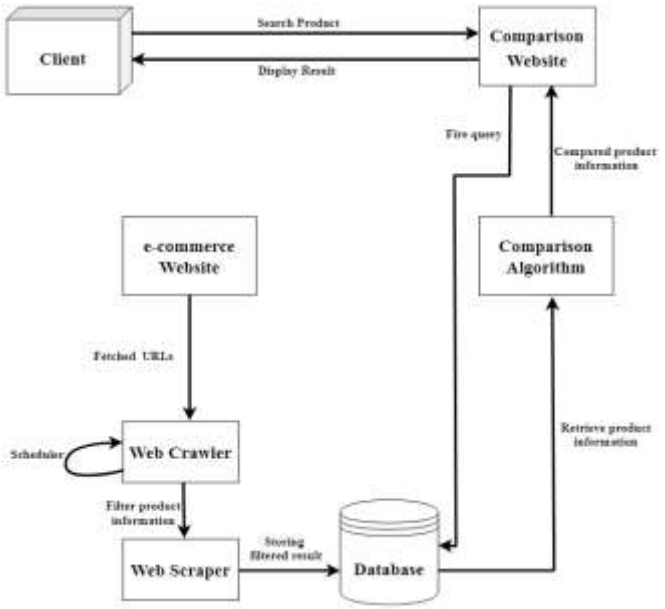
2. LITERATURE REVIEW

Web Crawler/scrapers were addressed for extracting URLs from different E-commerce websites. In reference paper focus was on product comparisons on behalf of humans. Also the focus was on implementing the architecture for online services by searching products from online websites and comparing the product amongst different websites and getting the cheapest price available on that product. The basic purpose was to search the product in cheapest price.

In Monitoring best product using Data Mining, the intention is to provide the customer a user experience that allows the customer to view and compare prices of a particular product from different websites and purchase the product which he/she finds suitable for him/her. This tends to reduce Time and effort put by customer providing customer ease and satisfactory results. It also tends to save customer from predatory pricing strategies imposed by different E-commerce websites. Eventually monitoring best product saves customers valuable time, efforts and money.

The technology being used is Python using Beautiful Soup and MongoDB for backend processing while Python Flask framework for front end.

3. SYSTEM ARCHITECTURE



E-commerce application have a certain model and standards that is followed in the industry. In Monitoring best Product using Data Mining technique, Intelligent agent is used to crawl through to different websites to fetch URL's of different products. The intelligent agent is a web crawler or a ShopBot that is an automated program that fetches the URL's from different e-commerce applications.

In most cases, users before purchasing the products online, they need to go different e-commerce applications to find the particular products at the cheapest price. Monitoring best product using Data Mining technique solves this problem of user by providing the user Products from different e-commerce applications at one place with different prices and schemes and offers that are offered by different e-commerce firms. This will provide the user privilege to choose products from different E-commerce applications which they consider is the best for them.

4. WORKING

An Intelligent agent is a web crawler running on the back-end of the website whereas front end technology provides a graphical user interface (GUI) for the users to communicate with the system. The explanation of the architecture is as follows: The Web-Crawler visits different e-commerce websites and fetch URLs from different e-commerce websites. The Filter performs filtration so as to remove useless URLs. Then the filtered URLs are stored in the local database. The database used is MongoDB. Web crawler periodically fetches the data from different e-commerce websites and if updates are available, then web crawler carries the updates and updates the local database. Whenever client searches for the products in the search bar of the comparative website, the local database is queried so as to retrieve the required results. The user can then compare products based on prices from different e-commerce websites. When user selects the best deal according to him and click on the buy button of the product then on clicking on the buy button, it triggers the user to original website to purchase the product.

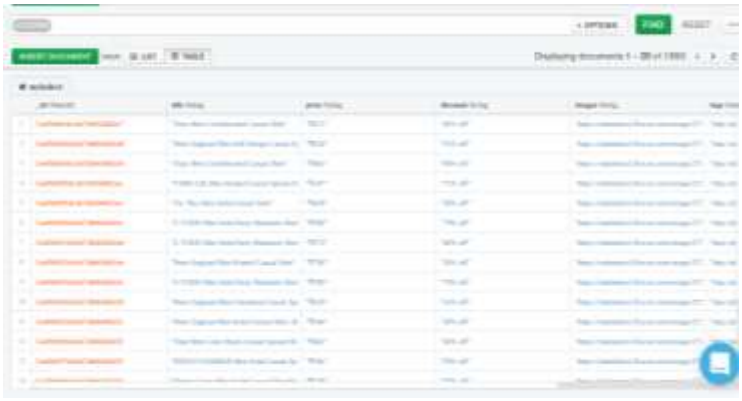


Fig -1: Web Scraping For Flipkart

5. CONCLUSIONS

The Monitoring best product using data mining technique is a price comparison engine that aims to facilitate the buyers to compare products from different e-commerce websites and purchase the product at the cheapest rate with best price. This way, the buyer has more power in his/her hands and can take better decision on different products at different price. Thus, this project saves buyers' efforts, time and money and also avoids user to physically visit each and everywhere e-commerce website. The beauty of this project is that it can be customize for a specific business segment and also can be used. It also helps the different Ecommerce applications to boost their business by providing them a platform to compete and do business in a more reasonable manner. By involvement of data intelligence, we can also comment and predict about the services and quality provided by the different e-commerce. So that user can be help to choose a better deal.

ACKNOWLEDGEMENT

We wish to express our sincere gratitude to Dr. Sanjay U.Bokade, Principal and Dr. Satish. Y. Ket , H.O.D of Department Computer Engineering of Rajiv Gandhi Institute of Technology for providing us an opportunity to do our project work on "Monitoring Best Product Using Data Mining Technique ".This project bears on imprint of many peoples. We sincerely thank our project guide Mr Dilip Dalgade for his guidance and encouragement in carrying out this synopsis work. Finally, we would like to thank our colleagues and friends who helped us in completing project work successfully.

REFERENCES

[1] "A Dive into Web Scraper World" Deepak Kumar Mahto, Member, IEEE, Lisha Singh, Member, IEEE, 2015.

[2] Yunhua Gu, Member, IEEE, Shu Shen, Member, IEEE, Jin Wang, Member, IEEE, Jeong-Uk kim, Member, IEEE, (2015). "Application of NoSQL Database MongoDB".

[3] Xue-mengLi, Member, IEEE, Yong-yi Wang, Member, IEEE, "Design and Implementation of an Indexing Method Based on Fields for Elasticsearch",2015.

[4] "Mining E-Commerce Data from E-Shop Websites" Andrea Horch, Member, IEEE, Holger Kett, Member, IEEE and Anette Weisbecker, Member, IEEE, 2015.

[5] "Comparison of E-commerce Products using web mining", Department of Computer Engineering, Sinhgad Institute of Technology and Science, Savitribai Phule Pune University

[6] Darshita Kalyani, Dr. Devarshi Mehta Gujarat, India. (June 2017) "Paper on Searching and Indexing Using Elasticsearch".

[7] "Implementation of Web Crawler", Pooja gupta, Member, IEEE, Kalpana Johari, Member, IEEE.