Project Weekly Progress Report  
Agile – Scrum

|  |  |
| --- | --- |
| Semester | W2023, SEM-2 |
| Course Code | AML-2404 |
| Section | Section 2 |
| Group Name | D |
| Student names/Student IDs | Jash Vaghasiya - C0884733  Nivedini Kathagonda - C0872720  Keval Parmar - C0882386  Monil Rupawala - C0882370  Sai Divya Madhuri Guntupalli - C0882360 |
| Reporting Week | Week-3 |
| Team Lead for the reporting week | Monil Rupawala |

## Progress Made in Reporting Week:

In the third week of our project, our team focused on the crucial task of collecting relevant data from various sources for the development of a precise and effective skincare recommendation system. One of our primary objectives was to scrape links from websites that provide skincare products. By extracting these links, we aimed to gather product-related details such as descriptions and prices. Additionally, we explored the possibility of utilizing external data platforms like Kaggle to supplement our dataset. This comprehensive approach to data gathering is essential in ensuring the accuracy and comprehensiveness of our skincare recommendation system.

In our endeavor to extract information from websites, we have embarked on a process commonly known as web scraping. Specifically, we have focused on scraping product details from a website, which involves retrieving information such as the product's description, price, features, and instructions on how to use it.

To achieve this, we have employed various libraries and tools that facilitate web scraping. One popular library frequently we used for scraping is Beautiful Soup. It provided a simple and intuitive interface for parsing HTML and XML documents, enabling us to extract desired elements from web pages. Additionally, we might use other libraries such as Requests, which allows us to send HTTP requests and retrieve the HTML content of a web page.

By combining the power of these libraries, we have been able to scrape the links of the products from the website. Subsequently, we extended our scraping process to extract more specific information related to each product. This includes paraphrasing and contextualizing details such as the product's description, price, features, and instructions, ensuring that the information is clear and comprehensible.

## Difficulties Encountered in Reporting Week:

One issue we are encountering is after successfully removing the HTML tags from the full product description obtained through web scraping, we are faced with the task of extracting separate sections of information within the text. These sections often include details such as the product's ingredients, instructions for usage, tips for optimal use, safety information, and various other relevant information.

Another issue we face is while scrapping the links of the site, whenever we click on any product link the location selection pop up shows up which is giving difficulty in extracting data. During the process of scraping product links from the website, we have encountered a challenge related to a location selection pop-up that appears when clicking on any product link. This pop-up poses difficulties in extracting the desired data. The issue arises because the location selection pop-up disrupts the automated scraping process by interrupting the flow of data extraction.