

## **Assignment: 9**

**AIM :** To create a review scrapper for any ecommerce website

### **PROBLEM STATEMENT /DEFINITION**

Create a review scrapper for any ecommerce website to fetch real time comments, reviews, ratings, comment tags, customer name using Python

### **OBJECTIVE**

- To understand the concept of Web scraping
- To understand the methods used in web scraping.

### **THEORY:**

Web scraping, web harvesting, or web data extraction is data scraping used for extracting data from websites. The web scraping software may directly access the World Wide Web using the Hypertext Transfer Protocol or a web browser. While web scraping can be done manually by a software user, the term typically refers to automated processes implemented using a bot or web crawler. It is a form of copying in which specific data is gathered and copied from the web, typically into a central local database or spreadsheet, for later retrieval or analysis. Web scraping a web page involves fetching it and extracting from it. [1]

Beautifulsoup4 library is used here to scrape the data.

The required Python libraries:

1. The request library to make network requests

To scrape data from a website, we need to extract the content of the webpage. Once the request is made to a website, the entire content of the webpage is available, and we can then evaluate the web content to extract data out from it. The content is made available in the form of plain text.

2. The html5lib library for parsing HTML

Once the content is available, we need to specify the library that represents the parsing logic for the text available. We'll be using the html5lib library to parse the text content to HTML DOM-based representation.

3. The beautifulsoup4 library for navigating the HTML tree structure

beautifulsoup4 takes the raw text content and parsing library as the input parameters. In our example, we have exposed html5lib as a parsing library. It can then be used to navigate and search for elements from the parsed HTML nodes. It can pull data out from the HTML nodes and extract/search required nodes from HTML structure.

#### Making the Request for the Web Content

Let's make the web request for the website to be scraped. We will be using the requests library. To start using the requests library, we need to install the third-party library using the following command

```
pip install requests
```

We will be scrapping the amazon e-commerce website for customer name,ratings and reviews. Let's first make a request to extract the content for the specified website. request.get makes a request to the webpage, which returns back the raw HTML content.

### **CONCLUSION:**

Beautiful Soup was used to scrape customer reviews from amazon Website.