Harnessing Ecommerce Reviews for Social Media Targeting

Anjali Hubli, Subhranil Swar, Divyansh Agarwal

[1]*Department of Electronics and Instrumentation Engineering, RVCE, Bangalore*
[2]*Department of Computer Science Engineering, RVCE, Bangalore*
[3]*Department of Electronics and Communication Engineering, RVCE, Bangalore*

Abstract - *Customer reviews from e-commerce platforms offer important insights into consumer behavior, preferences, and product reception in the modern digital marketplace. This study centers on utilizing these reviews for improving targeted advertising on social media platforms through a three-step process. Initially, we utilize web scraping methods to collect product reviews from online retail sites, extracting qualitative and quantitative information. Next, we examine the gathered information with a refined Large Language Model (LLM) like ChatGPT to understand target audience demographics, sentiment analysis, and important trends in consumer feedback. Ultimately, the information gathered from LLM analysis is utilized to create and launch extremely customized and focused advertising campaigns on social media, increasing interaction and sales of products. This article introduces a thorough model for combining e-commerce feedback with social media advertising, illustrating the impact of NLP-based tactics on improving digital marketing outcomes.*

Keywords—*Web scraping, LLM(Large Language Model),e-commerce, NLP(Natural Language Processing)*

1.INTRODUCTION

In the quickly changing digital market, feedback from consumers is essential for improving product offerings, developing marketing strategies, and increasing customer interaction. Online shopping websites such as Amazon, Flipkart, and others act as large collections of customer reviews, offering businesses valuable information on consumer preferences, issues, and feelings. Yet, the difficulty lies in successfully examining this disorganized data and converting it into useful insights that can support focused advertising efforts, particularly on social platforms.

This research paper suggests a new method that utilizes e-commerce reviews to improve targeting for social media advertising. Three main steps are followed by our methodology:

1. Gathering E-commerce Reviews via Web Scraping: Utilizing web scraping methods to collect reviews from e-commerce sites, obtaining both qualitative and quantitative feedback on a product.

2. Using a Large Language Model (LLM) for Review Analysis: After gathering reviews, a refined Large Language Model (LLM) like ChatGPT is used to assess customer demographics, preferences, and sentiment. This allows businesses to better pinpoint their target audience.

3. Utilizing insights from LLM analysis, targeted ads are run on social media platforms to tailor campaigns to potential customers' interests and behaviors, boosting conversion rates and product sales.

Through leveraging consumer feedback on e-commerce platforms, this method allows businesses to conduct data-centered social media campaigns that are more targeted, customized, and effective.

2. METHODOLOGY

Web scraping is an automated technique used to extract information from websites. It involves fetching a web page's content and parsing it to retrieve specific data, such as text, images, or metadata. This process can be accomplished through various tools and programming languages, with Python being one of the most popular due to its simplicity and rich ecosystem of libraries, including Selenium and BeautifulSoup.

2.1 Why Web Scraping?

While API integration is often the preferred method for data retrieval due to its structured and efficient access to information, there are specific scenarios where web scraping is more advantageous. Here's why web scraping was chosen for collecting reviews of the Bella Vita Luxury Woman Eau De Parfum Gift Set:

- Availability of Content: Many e-commerce platforms, including Amazon, may not provide public APIs that expose user-generated content like reviews. When APIs are not available or are limited in the data they offer, web scraping becomes a necessary alternative to access the required information.

- Comprehensive Data: Web scraping allows the collection of all available reviews, including those that may be filtered or restricted via APIs. For example, the API might return only a limited number of recent reviews, while web scraping can access the entire repository of customer feedback.

- User Interaction Simulation: Web scraping can mimic user behavior, enabling the retrieval of data that requires navigation through various sections of a website, such as clicking on "See all reviews" to access comprehensive reviews.
- Cost-Effective: For smaller projects or individual researchers, setting up an API integration might require significant development effort, including authentication mechanisms and adhering to rate limits. Web scraping can be implemented quickly with minimal overhead, especially for one-off data collection tasks.
- Granular Data Collection: Researchers can gather additional metadata (e.g., timestamps of reviews, reviewer usernames) that might not be included in API responses, providing richer context for analysis.

2.2 Transferring Data to LLM

A. Pre-trained LLM Approach (Using OpenAI's API):

1. Convert Data to Text: Prepare your data as a sequence of customer reviews in text format.

API Request to GPT-4: Send the review text to the OpenAI API and prompt the model to analyze the reviews to identify the target audience and generate ad ideas.

```
import openai

openai.api_key = "YOUR_API_KEY"

response = openai.Completion.create(

 engine="gpt-4",

 prompt=f"Analyze the following product reviews and identify the target audience. Also, generate ideas for ads and promotional campaigns:\n\n{reviews_text}",

 max_tokens=500,

 temperature=0.7

)

print(response['choices'][0]['text'])
```

A. Getting the Training Data Ready

Steps for Preparing Data:

1. Label Reviews:

Sentiment Labeling involves categorizing reviews by sentiment (positive, negative, neutral) to aid in interpreting customer satisfaction levels for the model.

Feature Tags: Attach labels to reviews for particular attributes or elements of a product (such as "cost", "performance", "strength", "styling"). This assists the model in concentrating on particular parts of the review. After preparing the data, you can adjust the LLM using resources such as OpenAI's fine-tuning API or Hugging Face's transformers library.

We have to use the following code with a prompt so that it can generate the promotional campaigns and target ads.

```
response = openai.Completion.create(

 model="fine-tuned-model-id",

 prompt="Analyze these product reviews for target audience insights and ad ideas:\n" + reviews_text,

 max_tokens=150   //Token means the number of words it generate based on the prompt often 150 is sufficient for a concise ad

)
```

Key Factors for Target Audience Analysis:

1. Demographic Insights: LLMs can analyze text data to infer demographic characteristics, even if not explicitly mentioned, based on writing style, review content, and preferences.

Prompts:

- "What are the common age groups and gender demographics based on the following reviews?"
- "Which customer types (e.g., students, professionals, seniors) does this product appeal to based on these reviews?"

2. Sentiment Analysis and Product Feedback: By performing sentiment analysis on the reviews, LLMs can categorize feedback into positive and negative segments and help identify customer pain points or praises.

Prompts:

- "What are the common likes and dislikes expressed in the following reviews?"
- "What aspects of the product do customers frequently mention, and are they positive or negative?"

3. Segmentation Based on Preferences: LLMs can help you segment customers into different groups based on product features they care about, such as price sensitivity, product quality, or specific use cases (e.g., travelers, tech enthusiasts, etc.).

Prompts:

- "Group the following reviews into segments based on the different types of customers mentioned, such as budget-conscious buyers or premium quality seekers."
- "Analyze these reviews and describe customer preferences for this product in terms of price, quality, and features."

4. Market Trends and Behavioral Patterns: LLMs can identify patterns in purchasing behavior or trends over time, such as seasonal preferences or shifting customer demands.

Prompts:

- "What seasonal trends or buying behaviors can you identify in the following customer reviews?"
- "What changes in customer feedback are noticeable over time in these reviews?"

2.3 Overview of the Methodology: CASE STUDY

To gather product reviews for the Bella Vita Luxury Woman Eau De Parfum Gift Set from the Amazon e-commerce site, this study uses a web scraping technique. Extracting qualitative and quantitative data for analysis, along with review text and related ratings, is the goal. The procedure makes use of pandas for data manipulation and Excel-formatted storage, as well as the Selenium library for online automation.

Libraries and Tools

Selenium: An open-source program for browser automation. It makes it possible to retrieve dynamic content from websites.

Pandas: A Python data analysis and manipulation toolkit that offers data structures like Data Frames to manage tabular data effectively.

WebDriver Manager: An application that automatically obtains the required Selenium drivers, simplifying the management of browser drivers.

Procedure for Gathering Data

Configuring the Environment

Setting up the required environment, including installing the essential libraries, is the first step in the data-gathering process. The pip commands listed below were used to do this:

party

Copy the code pip. Put Pandas in place Python XL Webdriver-manager for Selenium

Development of Scripts

A Python script called Amazon_Webscraping.py was created to automate the web scraping process. The stages the script takes are as follows:

Set up WebDriver: WebDriver Manager ensures the right driver is installed by enabling Chrome to initialize the Selenium WebDriver.

Getting to the Product Page: The following URL is used by the script to navigate to the Amazon product page:

This code can be copied: **https://amzn.in/d/aX3yYCG**

Page Loading: Time is used to introduce a temporary pause. To let the page load fully, use sleep().

To access the Reviews Section: The script navigates to the product page's reviews section by scrolling down. To view the full list of reviews, click the 'See all reviews' link, if it is available.

Extraction of Data

Until every review page is obtained, the fundamental data extraction procedure is carried out in a loop:

Review Gathering**:** The script uses particular CSS selectors to find the elements that hold the review text and star ratings for each review section:

To extract the review text, use:

Python

Code span**[data-hook='review-body']** should be copied.

The following methods are used to extract review ratings:

Code I **[data-hook='review-star-rating']** should be copied. Span

Storage of Data: A list of tuples containing the extracted reviews and ratings is kept for later processing.

Managing Pagination: The script looks for a 'Next' button so that it can move through the reviews' several pages until there are none left.

Storage of Data

Following the successful extraction of every review, the gathered data is transformed into a pandas DataFrame, which gives the data a structured framework. The following command is then used to save the data frame to an Excel file called webscraping.xlsx:

Code : **df.to_excel('C:/Users/subhr/Desktop/webscraping.xlsx', index=False)** should be copied.

The reviews and the ratings that go with them are tabulated in the Excel file that is created at the path that is supplied by this operation.

The approach outlined here provides a methodical way to use Pandas and Selenium for web scraping. When these tools are used together, product reviews from e-commerce platforms can be automatically extracted, allowing for data-driven analysis in marketing and consumer behavior research.

After scraping and storing the data, it must be readied prior to being input into an LLM for examination.

Preparing data by removing errors and inconsistencies before analyzing it.

1.Eliminate Repetitions: Make sure to remove any duplicates of reviews or entries.

2.Deal with Absent Information: Certain reviews might lack ratings or text content. These can be either deleted or dealt with using default values.

3.Tokenization and Text Preprocessing involve removing stopwords, cleaning special characters, and converting the text to lowercase if working with textual data. Nevertheless, advanced preprocessing is typically unnecessary for LLMs such as GPT.

import pandas as pd

 # Load the Excel file

df.pd.read_excel('C:/Users/subhr/Desktop/webscraping.xlsx')

 # Data cleaning: drop rows with missing values

df.dropna(subset=['Review Text', 'Star Rating'], inplace=True)

The above code will remove all the missing values like those reviews which are having only stars and no actual reviews and load the excel file for cleaning the file for smooth loading into LLM

After cleaning the data we have to load it into LLM. We can use Openai or similar llm models for loading with the help of following code

import openai

 # OpenAI API key setup

openai.api_key = 'your-openai-api-key'

 # Extract reviews as a single text block (you can limit the number of reviews for efficiency)

reviews_text = "\n".join(df['Review Text'].tolist()[:500]) # Limit to first 500 reviews for simplicity

 # LLM call for analysis

response = openai.Completion.create(

model="fine-tuned-model-id", # Use a fine-tuned or default model

prompt="Analyze these product reviews for target audience insights and ad ideas:\n" + reviews_text,

max_tokens=300

)

 # Display the result

print(response.choices[0].text.strip())

The above code will give the target audience,target ads and sentiment analysis(What percentage of crowd has positive and negative feeling about the product)

Once the above is executed, we might receive a result similar to this.

Insights about the intended audience:

1.Most of the clients are female in their mid-twenties to early forties.

2.Numerous reviews indicate that this product is commonly given as a gift.

3.Clients value the enduring fruity scent, perfect for career-oriented women or individuals who like going out at night.

Suggestions for advertising:

- "An Ideal Present for Women: A Deluxe Perfume Collection for Any Celebration."

- "A Perfume Set for the Modern Woman: Fruity, Long-Lasting, and Luxurious."

-"Act now to save 15% on your Bella Vita Luxury Set and enjoy long-lasting elegance with this exclusive deal."

Sentiment assessment of the review:

- The majority of reviews, totaling 85%, are positive, with a special mention of the fragrance and presentation of the gift packaging.

- Pricing issues mentioned in certain negative reviews may make it less attractive to younger customers with limited financial resources.

After extracting the insights, the LLM is able to create tailored advertisements using the extracted information.

Promotional campaigns should highlight the product as a high-end gift option, emphasizing its reputation for lasting scents.

Dealing with unfavorable feedback: In case some reviews point out the cost as a disadvantage, you could organize a temporary promotion to appeal to more price-sensitive clientele

3 .INFERENCE

The study describes a methodical process for gathering and evaluating customer feedback on the Bella Vita Luxury Woman Eau De Parfum Gift Set sold on the e-commerce platform Amazon. Utilizing Selenium for web scraping, Pandas for data manipulation, and OpenAI's GPT-4 for language processing allows for detailed analysis of consumer preferences, sentiments, and demographic trends. Preparing data involves conducting steps such as addressing duplicates, handling missing values, and performing basic tokenization to guarantee high-quality data for analysis. This allows for a precise LLM to produce in-depth analyses on customer emotions, target audience traits, and recommendations for personalized advertising approaches.

This research demonstrates the advantages of using both web scraping and LLMs for e-commerce companies to gain insight into customer views and actions. The approach allows for detailed analysis that can inform data-based marketing choices by employing sentiment labeling, feature tagging, and customer segmentation. Utilizing LLMs improves customer segmentation and offers personalized promotional concepts, benefiting brands in addressing feedback and issues, ultimately leading to campaign optimization and increased customer engagement.

4. CONCLUSIONS

The combination of web scraping and LLMs, as shown in this approach, offers an effective structure for gathering and examining customer feedback on e-commerce sites. The strategy makes good use of Selenium to browse through dynamic content, Pandas for organized data handling, and OpenAI's LLMs for producing insights that surpass traditional data analysis. This mix enables a comprehensive grasp of customer emotions and likes, which can be used directly for analyzing target audience and creating promotional strategies.

By utilizing fine-tuning and prompt engineering, companies can customize LLMs to offer precise information on customer demographics, preferences, and buying habits. This leads to improved customer segmentation, more polished ad material, and a better understanding of market trends. Moreover, improving customer satisfaction and retention can be achieved by responding to negative feedback with specific discounts or promotions determined by LLM analysis.

REFERENCES

1. Guo, X., Cheng, G., & Guo, H. (2021). "Fine-tuning pre-trained language models for consumer sentiment analysis." IEEE Access, 9, 40213-40224.

2. White, A., & Alam, A. (2020). "Utilizing Natural Language Processing for Targeted Marketing in E-commerce." Journal of Business Analytics, 5(3), 102-118.

3. Zhang, S., Xu, H., & Huang, L. (2021). "Sentiment Analysis in E-commerce with Pre-trained Language Models." Electronic Commerce Research and Applications, 40, 100958.

4. Yang, Z., Wang, Y., & Wei, C. (2022). "Fine-tuning and prompt engineering of language models for specific domains." Natural Language Processing Journal, 12(2), 229-243.

5. Qin, L., & Sun, C. (2022). "Targeted Advertising in E-commerce: A Language Model Approach." Marketing Science, 15(4), 301-318.

6. Hughes, J. P., & Alam, S. (2019). "Dynamic Web Content Scraping Using Selenium for Consumer Review Analysis." Journal of Applied Analytics, 22(5), 471-482.

7. Chen, X., Li, J., & Zhang, T. (2021). "Customer Segmentation using Machine Learning and NLP in E-commerce." Journal of Retailing and Consumer Services, 61, 102563.

8. Gupta, P., & Patel, R. (2020). "E-commerce Data Mining: Techniques for Understanding Customer Reviews." Data Science Journal, 19(1), 1-12.

9. Bian, J., & Yang, F. (2021). "Applications of Language Models in E-commerce Consumer Analysis." IEEE Transactions on Consumer Electronics, 67(2), 1027-1035.

10. Cui, H., & Wang, M. (2021). "Understanding Customer Behavior in E-commerce Using NLP and Machine Learning." Journal of Consumer Research, 45(6), 1256-1273.

11. Li, X., & Cao, W. (2021). "Enhancing Consumer Experience with Fine-tuned Language Models." International Journal of Market Research, 63(4), 442-459.

12. Kaushik, S., & Desai, A. (2019). "Dynamic Content Extraction with Selenium for Marketing Applications." Information Systems Research, 30(3), 785-800.

13. Tang, J., & Li, L. (2021). "Target Audience Analysis with Deep Learning in E-commerce." Journal of Marketing Research, 58(2), 327-338.

14. Zhou, L., & Xu, Y. (2021). "Customer Feedback Analysis in E-commerce: A Case Study Using GPT Models." Journal of Retailing, 97(2), 151-168.

15. Kim, Y., & Han, S. (2020). "Understanding Demographic Trends in Online Reviews with NLP." Journal of E-commerce Research, 10(1), 39-51.

16. Wang, J., & Zhao, P. (2020). "Sentiment Labeling for E-commerce Reviews Using AI Models." AI and Big Data in Business Journal, 15(3), 215-227.

17. Sun, Y., & Lin, W. (2021). "Exploring Consumer Preferences with Fine-tuned AI in E-commerce." Electronic Markets, 31(4), 691-702.

18. Liu, H., & Qian, H. (2021). "NLP Techniques for Product Feedback Analysis in E-commerce." Big Data Research, 23, 100190.

19. Rao, Y., & Zhang, M. (2022). "Data-Driven Consumer Insights Using GPT for Retail Strategy." Journal of Interactive Marketing, 61, 27-39.

20. Wilson, J., & Scott, T. (2021). "Analyzing E-commerce Reviews with Transformer Models." Journal of Marketing Science, 12(4), 505-517.

21. Figueroa, A., & Zhou, H. (2020). "Machine Learning for Consumer Behavior Prediction in E-commerce." IEEE Transactions on Neural Networks and Learning Systems, 31(11), 4451-4463.

22. Nguyen, T., & Clark, P. (2021). "Transforming E-commerce Insights with Language Models." Journal of Consumer Analytics, 3(3), 145-157.

23. Brown, R., & Miller, S. (2022). "Market Trend Analysis Using Transformer-based Models." Data Science in Business Journal, 8(2), 215-225.

24. Cai, S., & Jiang, W. (2020). "Leveraging AI for Targeted Advertising in Retail." Journal of Retailing and Consumer Services, 55, 102087.

25. Zhang, K., & Wang, L. (2021). "Impact of AI on E-commerce Consumer Behavior Analysis." Computers in Human Behavior, 124, 106911.