

# Amazon Customer Review Summary Using Web Scraping and Artificial Intelligence Techniques

## Introduction

In an increasingly digital age focused on online sales, customers frequently share their opinions on purchased products or services. These reviews are of significant interest to potential buyers, who often trust the advice of other users more than traditional advertising.

However, reading all reviews to obtain a comprehensive picture is time-consuming, necessitating an effective summary of a large volume of opinions. This summary should highlight the strengths and weaknesses of the product, providing a valuable service to both users, for making informed decisions, and companies, for improving products, increasing customer satisfaction, and driving business growth.

To achieve this goal, web scraping techniques are used to extract information from online sales sites, and advanced text processing techniques are employed to transform texts into analysable data, allowing for the understanding of the "sentiment" towards a certain product or service.

With the advent of generative Artificial Intelligence (AI), such as ChatGPT, this process, which previously required specialized software and highly qualified personnel, can now be managed by users with less expertise, as demonstrated in this study.

## Study objective

The objective of the study was to extract and analyze all reviews provided by buyers of a book sold on Amazon Italy (<https://www.amazon.it/cimitero-Praga-Umberto-Eco/dp/8845278646/>) using a generative AI model to evaluate the opinions and provide a brief summary of the positive and negative reviews.



Figure 1 – The main product page

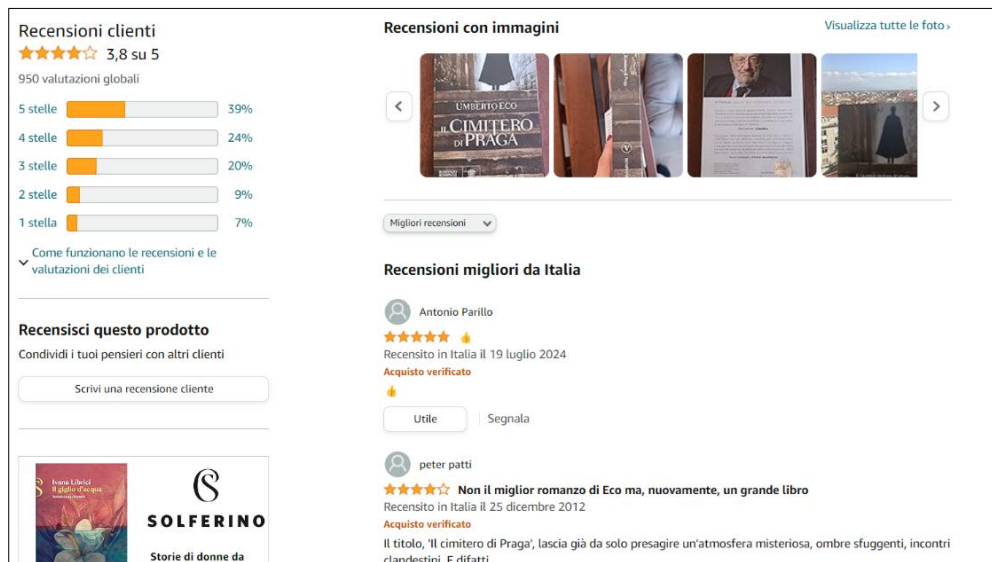


Figure 2 – The reviews page

## Methodology

- Access to the product page on Amazon and **extraction of relevant fields from the HTML** source code: These fields populate the columns of a data table where each row corresponds to a review.
- **Reviewer's name**
- **Rating** (in "stars", from 1 to 5)
- **Review text**

This was accomplished through a **Python program**, with minimal corrections, by **GPT-4o** (thanks OpenAI!) by providing only the link to the page of interest.

```

1  import requests
2  from bs4 import BeautifulSoup
3  import pandas as pd
4  import os
5  import time
6
7  custom_headers = {
8      'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/111.0.0.0 Safari/537.36'
9  }
10
11 def get_soup(url):
12     response = requests.get(url, headers=custom_headers)
13     if response.status_code != 200:
14         print("Error in getting webpage: ", response.status_code)
15         exit(-1)
16     soup = BeautifulSoup(response.text, "lxml")
17     return soup
18
19 def get_reviews(soup):
20     review_elements = soup.select("div[data-hook='review']")
21     scraped_reviews = []
22     for review in review_elements:
23         r_author_element = review.select_one("span[data-hook='review-author']")
24         r_author = r_author_element.text if r_author_element else None
25
26         r_rating_element = review.select_one("span[data-hook='review-star-rating']")
27         r_rating = r_rating_element.text.replace(" su 5 stelle", "") if r_rating_element else None
28
29         r_content_element = review.select_one("span[data-hook='review-body']")
30         r_content_full_element = review.select_one("span[data-hook='review-body'] span")
31         if r_content_full_element:
32             r_content = r_content_full_element.text.strip()
33         else:
34             r_content = r_content_element.text.strip() if r_content_element else None
35
36         r = {
37             "author": r_author,
38             "rating": r_rating,
39             "content": r_content
40         }
41         scraped_reviews.append(r)
42     return scraped_reviews
43

```

Figure 3 – Some lines of Python code for web scraping

- Adding an additional column to the table containing the prompt that combines the question text with the review text:

"L'opinione in questo testo è positiva, neutrale o negativa?  
 Rispondi con uno di questi termini: 'Positiva', 'Neutra',  
 'Negativa'. Questo è il testo da analizzare: " + <review text>

| S ▲ Recensore       | S Valutazione | S Recensione          | S Prompt                                                                                                                                                                                                                       |
|---------------------|---------------|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ALIGGIE             | 5 stelle      | bello                 | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: bello                                                       |
| Abbo                | 2 stelle      | Mi aspettavo qu...    | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: Mi aspettavo qualcosa di diverso e qualcosa di più.P.       |
| Alalea              | 5 stelle      | Mi associo ad alt...  | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: Mi associo ad altre recensioni che segnalano una diff       |
| Alan L.             | 5 stelle      | Come tutte le o...    | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: Come tutte le opere narrative di Eco c'è da imparare        |
| Aldo                | 1 stella      | Capitano libri noi... | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: Capitano libri noiosi, ma di valore. Questo e' noioso .     |
| Aldo                | 2 stelle      | Gran bell'inizio q... | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: Gran bell'inizio quello di questo libro di Eco. Come del    |
| Ale & Cri           | 1 stella      | noioso oltre ogni...  | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: noioso oltre ogni umana comprensione e non è il dife        |
| Alessandro Drovandi | 4 stelle      | Bella lettura... s... | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: Bella lettura... scegliere Eco non è un azzardo... leg      |
| Alessio             | 1 stella      | Libro interesstan...  | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: Libro interessante ed istruttivo dal punto di vista sto     |
| Amazon Customer     | 2 stelle      | L'ironia delle des... | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: L'ironia delle descrizioni, l'intreccio psicologico e l'amb |
| Andrea              | 4 stelle      | L'ho trovato acc...   | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: L'ho trovato accattivante con il finto 'sdoppiamento'       |
| Andreasasa          | 5 stelle      | Al di là d'imprev...  | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: Al di là d'imprevedibili e caratteristiche ECOtrame, in     |
| Angelo              | 2 stelle      | Lasciato dopo q...    | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: Lasciato dopo qualche capitolo...non mi ha mai pres         |
| Angelo Partoini     | 5 stelle      | Ottimo acquisto...    | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: Ottimo acquisto. Grande acquisto di un grande autor         |
| Anna                | 5 stelle      | Diverentissimo l...   | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: Diverentissimo libro eccezionale, Umberto Eco è un          |
| Antelope Cobbler    | 4 stelle      | In un epoca com...    | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: In un epoca come la nostra in cui il cospirazionismo t      |
| Antonella Alberti   | 3 stelle      | Il romanzo proc...    | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: Il romanzo procede un po' a rilento, rispetto alle avv      |
| Antonio             | 4 stelle      | Fondamentalm...       | L'opinione in questo testo è positiva, neutrale o negativa? Rispondi con uno di questi termini: 'Positiva', 'Neutra', 'Negativa'. Questo è il testo da analizzare: Fondamentalmente a un lettore avvezzo ai particola          |

Figure 4 – The input table to the LLM

- Connection to OpenAI's Large Language Model (LLM)**, which uses AI for text understanding and generation.
- Sending the table containing prompts to the OpenAI server.** The LLM will process these prompts and provide responses based on the review text.
- Validation of the results** by comparing the ratings provided by reviewers and the sentiment detected by AI. As shown in Figure 5, there is a strong association between the respective values.

| Valutazioni dei recensori e "sentiment" da ChatGPT |          |        |          |
|----------------------------------------------------|----------|--------|----------|
| Valutazione                                        | Negativa | Neutra | Positiva |
| 1 stella                                           | 27       | 2      | 0        |
| 2 stelle                                           | 21       | 4      | 0        |
| 3 stelle                                           | 7        | 16     | 2        |
| 4 stelle                                           | 1        | 14     | 31       |
| 5 stelle                                           | 3        | 7      | 58       |

Figure 5 – The cross table comparing ratings and sentiment

- **A new use of the LLM to create two short summaries:** one by aggregating all reviews with a “positive” sentiment and the other by aggregating those with a “negative” sentiment.

| Riassunto dei "sentiment" positivi                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Riassunto dei "sentiment" negativi                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Il libro "Il cimitero di Praga" di Umberto Eco racconta la storia di Simonini, un protagonista senza scrupoli che partecipa ad alcuni degli avvenimenti più importanti dell'Europa della seconda metà dell'ottocento. Simonini è un falsario che, con la sua doppia personalità, inventa i "Protocolli dei Savi di Sion", un falso che diventa la base per i movimenti antisemiti successivi. Il romanzo è una satira storico-politica che mette in luce la tendenza umana a creare stereotipi e pregiudizi etnici. Eco, con la sua prosa colta e ironica, ci mostra la follia dell'odio verso lo straniero e il diverso. Il libro è una lettura avvincente, anche se meno coinvolgente del precedente "Il nome della rosa", e sconsigliato a chi non ama una letteratura un po' erudita.</p> | <p>Il romanzo di Umberto Eco, "Il cimitero di Praga", inizia promettente con una narrazione dei fatti dell'unità d'Italia e della storia di Ippolito Nievo, ma poi diventa confuso e frettoloso. Il protagonista, con problemi psichici, dopo la morte dei parenti si dedica a imbrogliare il prossimo. La storia declina man mano che si procede, con temi come complotti, massoni, gesuiti e spie che si sovrappongono senza una trama ben definita. L'autore pesca da molte fonti, ma il risultato è noioso e ripetitivo. Nonostante la vasta erudizione di Eco, il libro risulta sgradevole e poco coinvolgente. Non si consiglia la lettura, soprattutto per il rischio di alimentare idee sbagliate. Ci sono molte altre opere letterarie più interessanti e piacevoli da leggere.</p> |

Figure 6 – The summary of "positive" and "negative" reviews

- Assembling all the results into a single view to have a complete analysis overview.

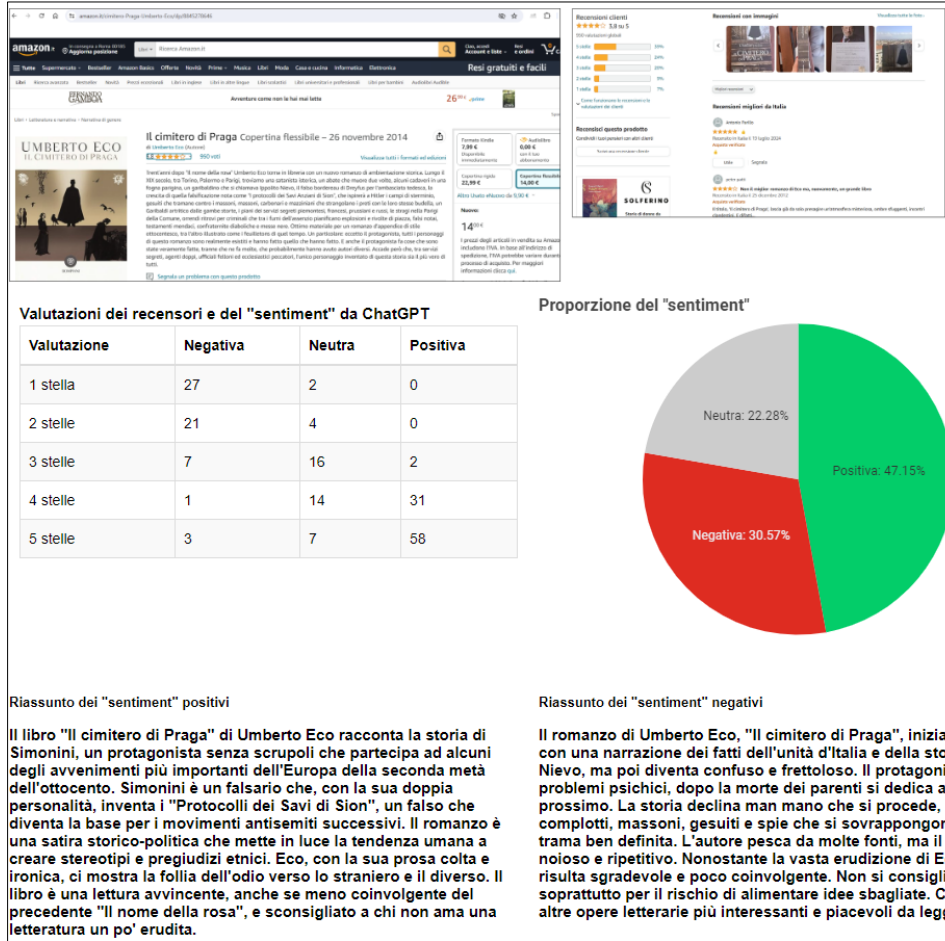


Figure 7

Overview of the various produced outputs

"This is the general summary provided by Amazon, generated by AI from customer reviews:

*“Readers say the book is interesting for its historical setting, which is very relevant given the recurrence of the events. They also appreciate the author's engaging style and the quality of the product. Overall, they consider it a good purchase and recommend it. However, some find the narrative slow and inconclusive.”*

## Tools used

**LLM** (specifically the **GPT-3.5-turbo-instruct** model) from OpenAI through API Key for authentication and authorization to use the model.

**Python** version 3.6 for web scraping.

**KNIME Analytical Platform** version 5.3 for the user interface and integration of various functionalities:

- Authentication and access to OpenAI's LLM
- Integration of a Python script for web scraping
- Creation of prompts for querying the LLM
- Visualization of results

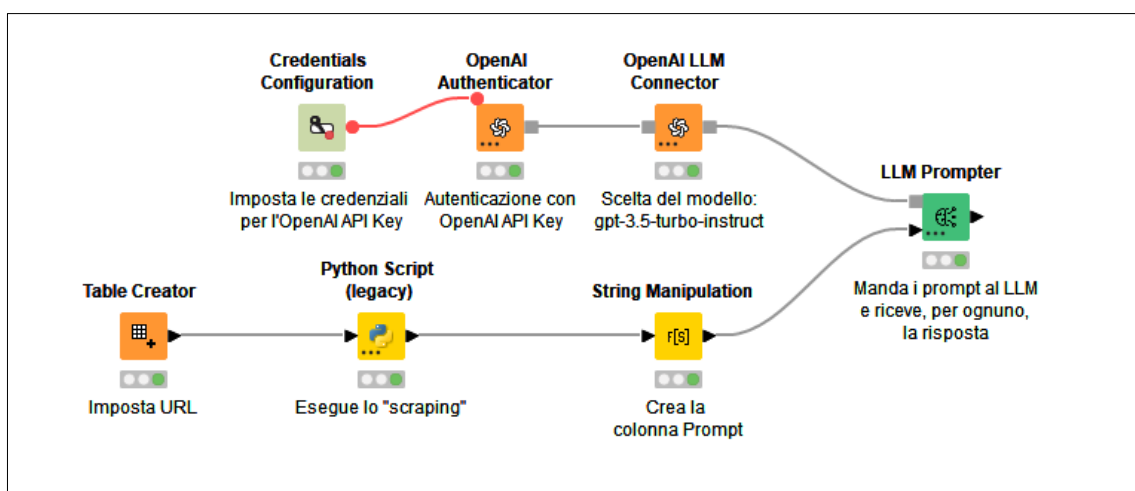


Figure 8 – Partial view of the KNIME workflow

## Conclusion

This study presented a process for extracting and analysing textual data regarding product reviews on an online sales site.

After the web scraping phase, sentiments were synthesized using generative AI, thus saving the long and laborious task of text processing and analysis, typically assigned to expert personnel.

By utilizing a user interface, the various steps were arranged in a workflow through functional nodes, optimizing and automating the extraction and analysis process.

This approach not only provides a useful service to customers but is also crucial for companies to improve their products and marketing strategies.