

Riassunto delle recensioni dei clienti su Amazon utilizzando tecniche di web scraping e Intelligenza Artificiale

Introduzione

In un'epoca sempre più digitalizzata e orientata verso le vendite online, i clienti condividono spesso le proprie opinioni sui prodotti o servizi acquistati. Queste recensioni suscitano l'interesse degli acquirenti, che spesso si fidano più dei consigli degli altri utenti rispetto alla pubblicità tradizionale.

Tuttavia, leggere tutte le recensioni per ottenere un quadro completo richiede tempo, rendendo necessaria una sintesi efficace di una grande quantità di opinioni. Questa sintesi dovrebbe evidenziare i punti di forza e di debolezza del prodotto, fornendo un servizio utile sia agli utenti per prendere decisioni informate, sia alle aziende per migliorare i prodotti, aumentare la soddisfazione dei clienti e stimolare la crescita del business.

Per raggiungere questo obiettivo, si utilizzano tecniche di “web scraping” per estrarre informazioni dai siti di vendita online e tecniche avanzate di text processing per trasformare i testi in dati analizzabili, permettendo di comprendere il “sentiment”¹ verso un certo prodotto o servizio.

Con l'avvento dell'Intelligenza Artificiale (IA) generativa, come ChatGPT, questo processo, che in passato richiedeva software specializzato e personale altamente qualificato, può essere gestito anche da utenti non particolarmente esperti, come dimostrato nel presente studio.

Scopo dello studio

Lo scopo dello studio è stato quello di estrarre e analizzare tutte le recensioni fornite dagli acquirenti di un libro venduto su Amazon Italia (<https://www.amazon.it/cimitero-Praga-Umberto-Eco/dp/8845278646/>) utilizzando un modello di IA generativa per valutare le opinioni e fornire un breve riassunto delle recensioni positive e negative.

¹ La Sentiment Analysis (analisi del “sentiment”) è un processo che utilizza l'elaborazione del linguaggio naturale per determinare l'emozione o l'opinione espressa in un testo.



Figura 1 – La pagina principale del prodotto

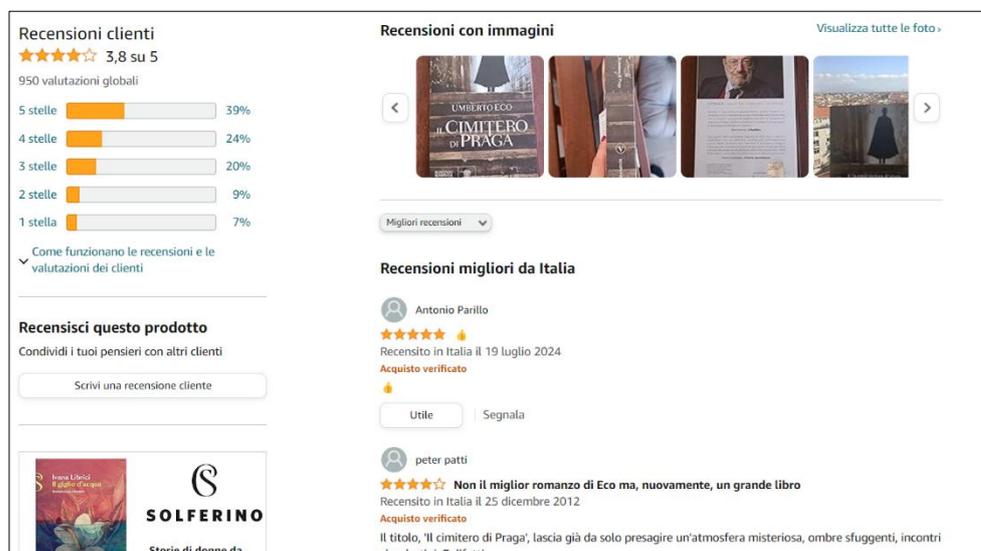


Figura 2 – La pagina delle recensioni

Metodologia adottata

- Accesso alla pagina del prodotto venduto su Amazon ed estrazione dei campi di interesse **dal codice sorgente HTML**. Questi campi alimentano le colonne di una tabella di dati dove ogni riga corrisponde a una recensione.
 - **Il nome del recensore**
 - **La valutazione** (in “stelle”, da 1 a 5)
 - **Il testo della recensione**

Questo è stato possibile grazie a un **programma sviluppato da GPT-4o** (grazie OpenAI!) **in linguaggio Python**, apportando minime correzioni, fornendo semplicemente il link della pagina.

```

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.a-profile-name")
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

```

Figura 3 –Alcune righe di codice in linguaggio Python per il web scraping sviluppato da GPT-4

- Inserimento nella tabella di una colonna aggiuntiva contenente il testo in ingresso (detto “prompt”) che combina il testo della domanda con il testo della recensione:

"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: " + <testo della recensione>

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
Rbbo	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 interessan...	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 press
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	Divertentissimo 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: Divertentissimo 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 parcol

Figura 4 – La tabella di input al LLM

- **Connessione al Large Language Model (LLM)² di OpenAI** per la comprensione e la generazione delle risposte a ogni "prompt".
- **Invio della tabella contenente i prompt al server di OpenAI.** Il modello LLM elaborerà questi prompt e **fornirà le risposte** basate sul testo delle recensioni.
- **Validazione dei risultati** ottenuta confrontando le valutazioni fornite dai recensori e i “sentiment” rilevati dall'IA. Come si evince dalla figura 5, c'è una forte associazione tra i rispettivi valori.

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

Figura 5 – La tabella incrociata che mette a confronto le valutazioni e il “sentiment”

- **Nuovo utilizzo del LLM per creare due brevi riassunti:** il primo aggrega tutte le recensioni con “sentiment” positivo, il secondo aggrega quelle con “sentiment” negativo.

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>

Figura 6 – Il riassunto delle recensioni “positive” e di quelle “negative”

Questa la sintesi generale che fornisce Amazon, sempre generata dall'IA. a partire dal testo delle recensioni dei clienti:

“I lettori affermano che il libro è interessante per la sua ambientazione storica, molto attuale vista la ricorrenza. Apprezzano inoltre l'aspetto accattivante dell'autore e la qualità del prodotto. Nel complesso, lo considerano un buon acquisto e consigliato. Tuttavia, alcuni trovano la narrazione lenta e inconcludente.”

² Un modello linguistico di grandi dimensioni è un **IA addestrata su grandi quantità di testo** per la comprensione e la generazione di un testo in linguaggio naturale.

- Integrazione di tutti i risultati ottenuti in un'unica vista per avere un quadro completo dell'analisi.

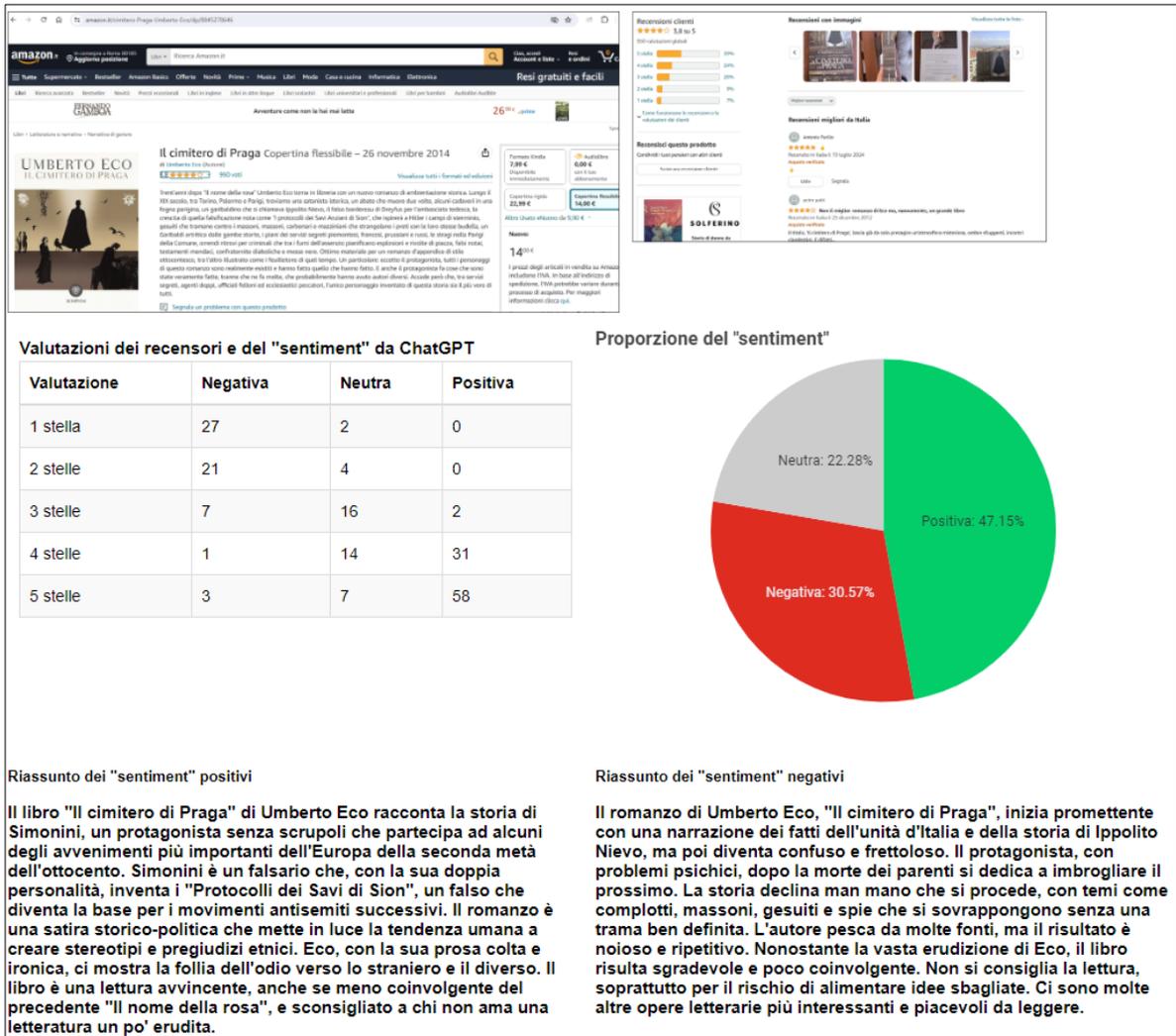


Figura 7 – Vista d'insieme dei vari output prodotti

Strumenti Utilizzati

LLM (in particolare il modello **GPT-3.5-turbo-instruct**) di OpenAI attraverso API Key per l'autenticazione e l'autorizzazione all'utilizzo del modello.

Python versione 3.6 per il Web Scraping.

KNIME Analytical Platform versione 5.3 per l'interfaccia utente e l'integrazione delle varie funzionalità:

- Autenticazione e accesso al modello LLM di OpenAI
- Integrazione di uno script Python per il web scraping
- Creazione dei prompt per le interrogazioni al modello LLM
- Visualizzazione dei risultati

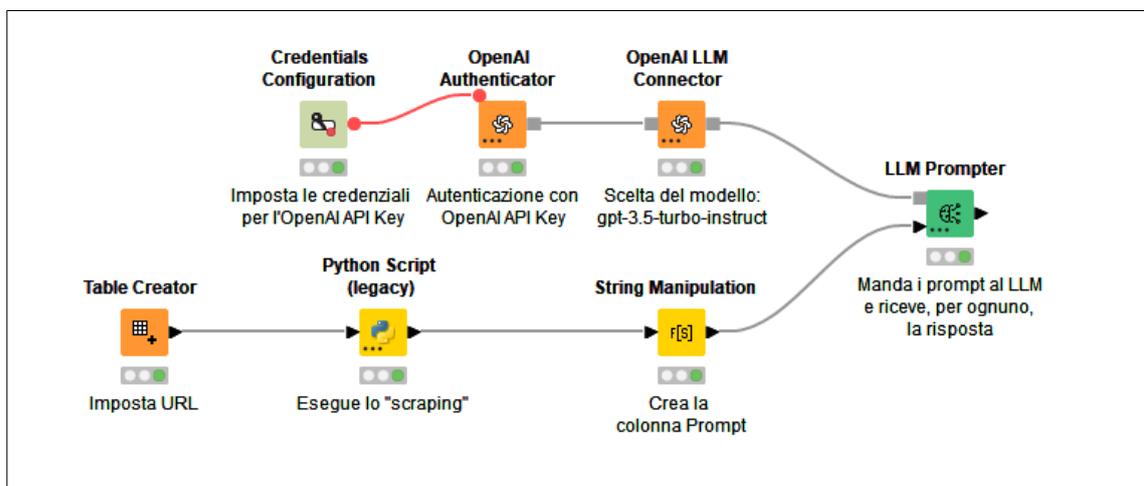


Figura 8 – La pagina delle recensioni

Conclusione

Questo studio ha presentato un processo di estrazione e analisi dei dati testuali riguardanti le recensioni di un prodotto su un sito di vendite online.

Dopo la fase di web scraping, si sono sintetizzati i "sentiment" attraverso l'utilizzo dell'IA generativa, risparmiando così il lungo e faticoso lavoro di elaborazione del testo e analisi, solitamente affidato a personale esperto.

Grazie all'utilizzo di un'interfaccia utente, i vari passaggi sono stati disposti in un flusso di lavoro attraverso nodi funzionali, ottimizzando e automatizzando il processo di estrazione e analisi.

Questo approccio non solo fornisce un servizio utile ai clienti, ma è anche fondamentale per le aziende per migliorare i loro prodotti e strategie di marketing.