Sentiment Analysis in Hotel Recommendation Systems

Leveraging Sentiment Analysis to Personalize Hotel Recommendations

Joel Luis and Swizel Monteiro

2024-09-17

Abstract

With the vast amount of hotel information available on the web and through social media, selecting the ideal accommodation has become increasingly complex for travelers. Hotel recommendation systems aim to simplify this process by offering personalized suggestions.

This project presents a hotel recommendation system that leverages sentiment analysis to enhance recommendation accuracy. By analyzing user reviews collected from Google accommodations data, the system extracts and interprets users' preferences and sentiments. Reviews are preprocessed, analyzed for sentiment, and used to gauge user preferences. The system then matches these preferences with hotel features such as amenities and descriptions to generate tailored recommendations. Additionally, the system incorporates budget constraints and user-selected amenities to refine the suggestions further.

The effectiveness of the system is evaluated using real-world data, demonstrating improved recommendation precision compared to traditional methods. This approach not only streamlines the hotel selection process but also offers a more personalized and satisfying user experience.

Table of Contents

1 Introduction

The proliferation of online platforms and social media has led to an exponential growth in user-generated content, including reviews and ratings of hotels and (Jardim and Mora, 2022). Travelers increasingly rely on this wealth of information to make informed decisions about their lodging options, seeking recommendations that align with their preferences and expectations (Abbasi-Moud et al., 2019). However, the sheer volume of data and the diverse nature of user reviews present a significant challenge for travelers who must sift through vast amounts of information manually (Jones, Clark, and Martin, 2018).

To address this challenge, various hotel recommendation systems have been developed to provide personalized suggestions based on user preferences(Jardim and Mora, 2022) . These systems typically analyze user input and historical data to recommend hotels that match users' criteria. Some systems focus on clustering users with similar profiles and recommending hotels based on collective preferences (Abbasi-Moud et al., 2021). For instance, users with similar travel histories might receive identical recommendations, which can be effective but lacks nuance.

Another approach involves analyzing user reviews to extract preferences and match them with hotel features. This method leverages textual analysis to identify common themes and preferences . However, many systems that analyze reviews focus primarily on keyword frequency, potentially overlooking the importance of sentiment expressed in the reviews. Negative sentiments or critical feedback might be misinterpreted as preferences if not properly addressed.(Bui, 2020))

Sentiment analysis, a technique used to determine the emotional tone of a text, is a crucial but often underutilized aspect of recommendation systems in the hospitality industry (Ray et al., 2021)). A comprehensive hotel recommendation system should not only identify user preferences through semantic analysis but also integrate sentiment analysis to differentiate between positive and negative feedback accurately. Additionally, incorporating contextual information, such as budget constraints and specific amenities, can further enhance the relevance of recommendations.

This project introduces an advanced hotel recommendation system that integrates sentiment analysis to refine recommendations based on user reviews. The system processes textual data to extract sentiment scores, matches these scores with hotel features, and filters recommendations based on user preferences, amenities, and budget constraints. The proposed system is developed using Python and tested on a dataset gathered from

Google accommodations data. Preliminary results demonstrate the system's ability to provide more accurate and personalized hotel recommendations compared to traditional methods.

2 Proposed Recommendation system

The proposed hotel recommendation system enhances the accuracy and relevance of hotel suggestions by incorporating sentiment analysis into the recommendation process. The system integrates several key components to achieve this objective:

2.1 Overview

The recommendation system is developed using Flask for the web interface, Pandas for data manipulation, and TextBlob for sentiment analysis. It processes user reviews to extract preferences and generate personalized hotel recommendations. The system's architecture is designed to efficiently handle user inputs, data processing, and recommendation generation.

2.2 Data Sources

The system leverages web scraped data from Google accommodations to provide comprehensive insights for hotel recommendations. This dataset encompasses several key elements: hotel descriptions, amenities, prices, and user reviews, along with review dates.

The descriptions offer detailed information about each hotel's features, while the amenities list includes various facilities such as laundry, pools, gyms, spas, and restaurants. Prices reflect the cost of staying at each hotel per night, and user reviews provide valuable feedback for sentiment analysis.

The dataset is structured in a CSV file and organized into a DataFrame with columns for Hotel, Description, Review, ReviewDate, Amenities, and Price. This structured format facilitates efficient data manipulation and analysis, enabling the system to accurately extract user preferences and match them with hotel features to generate relevant recommendations.

2.3 Sentiment Analysis

Sentiment analysis in the proposed system is conducted using TextBlob, a Python library specifically designed for processing textual data. TextBlob provides straightforward tools to perform various natural language processing tasks, including sentiment analysis. By analyzing user reviews, TextBlob calculates sentiment scores that capture the emotional tone of the comments. This involves determining whether the sentiments expressed in the reviews are positive, negative, or neutral, allowing the system to gauge overall user satisfaction with each hotel.

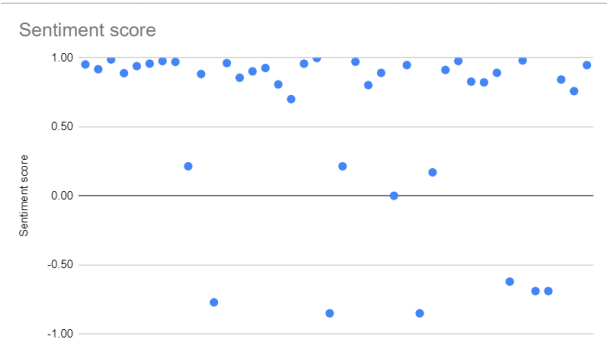


Figure 1: Variation in sentiments of different hotels

2.3.1 Sentiment Score Calculation

The sentiment score for each review is computed based on its polarity, which ranges from -1 (most negative) to +1 (most positive). This polarity score reflects the review's emotional tone, with positive scores indicating favorable sentiments and negative scores representing dissatisfaction. Hotels are then ranked according to their average sentiment scores, with higher scores signifying more positive reviews. This ranking process enables the system to prioritize hotels with higher user satisfaction, ensuring that recommendations are based on the most favorable feedback.

Top Hotel Recommendations Based on Your Selected Services

	Hotel	Description	Review	ReviewDate	Amenities	Price	RecencyWeight	SentimentScore	AdjustedSentiment
0	Eco Resort Bali	eco-friendly, adventure, beach	Amazing service, clean rooms, loved the adventure activities!	2024-09-10	[laundry, pool, gym, spa, restaurant]	100	0.983562	0.8	0.786849

Back to Home

Figure 2: Recommendation based on sentiment score and related preferences

2.4 User Preferences and Matching

2.4.1 Capturing User Input

Users interact with the recommendation system via a web form designed to capture their preferences. Through this form, users can specify their desired services, such as "Clean Room," "Good Location," and "Great Service," as well as select specific amenities like laundry, TV, pool, gym, spa, and restaurant. Additionally, users set their budget constraints to filter hotels that fit within their financial range. This collected data forms the basis for personalized recommendations, ensuring that the suggestions provided align closely with individual user preferences and constraints.

2.4.2 Matching Preferences

To match user preferences with hotel options, the system employs keyword-based filtering techniques. This process involves comparing the selected preferences and amenities with the descriptions of available hotels. Hotels that contain keywords and attributes matching the user's input are initially identified. The relevance of these matches is then refined through additional criteria, including sentiment scores derived from user reviews. This refinement process ensures that recommended hotels not only meet the specified preferences but also have positive user feedback, thereby enhancing the overall recommendation quality. The final recommendations are presented to users based on the alignment of their preferences, hotel descriptions, and sentiment scores, ensuring a highly personalized and satisfying selection.

2.5 Filtering and Recommendations

2.5.1 Amenities and Budget Filtering

The recommendation system filters hotels based on user-selected amenities and budget constraints. Users specify the amenities they desire, such as a pool, gym, or restaurant, and set their budget limits. The system then eliminates hotels that do not meet these criteria, ensuring that only those which fulfill the specified requirements are considered for recommendation. This step guarantees that the recommendations are tailored to the user's specific needs and financial constraints.

2.5.2 Recommendation Generation

After filtering, the remaining hotels are ranked according to their sentiment scores, which reflect the overall user satisfaction from reviews. The system prioritizes hotels with higher sentiment scores, indicating more favorable reviews and positive experiences from previous guests. This ranking ensures that the top recommendations are those with the best-reviewed experiences, thereby enhancing the relevance and quality of the suggestions provided to the user.

2.6 Recency Integration

2.6.1 Review Recency

To ensure that recommendations are based on current and relevant information, the system incorporates the recency of user reviews. Recent reviews are given more weight in the sentiment analysis, as they provide a more accurate reflection of the hotel's current quality and conditions. This approach helps to filter out outdated information and focus on the latest guest experiences.

2.6.2 Timely Recommendations

By integrating recency into the recommendation process, the system ensures that the suggestions are up-to-date and relevant. This means that users receive recommendations based on the most recent feedback and conditions, making the suggestions more applicable to current scenarios and improving overall recommendation accuracy.

2.7 User Interface

2.7.1 Web Interface

The user interface for the recommendation system is developed using Flask and Material-UI, which provide a modern, responsive, and intuitive experience. Users can easily input their preferences, view recommendations, and navigate the system through a well-designed web interface that enhances usability and accessibility.

2.7.2 Recommendations Display

Recommended hotels are presented with essential details, including the hotel name, description, amenities, and price. This format ensures that users can make informed deci-

sions based on comprehensive information about each hotel. The clear and concise display of relevant data helps users quickly evaluate their options and select the best fit for their needs.

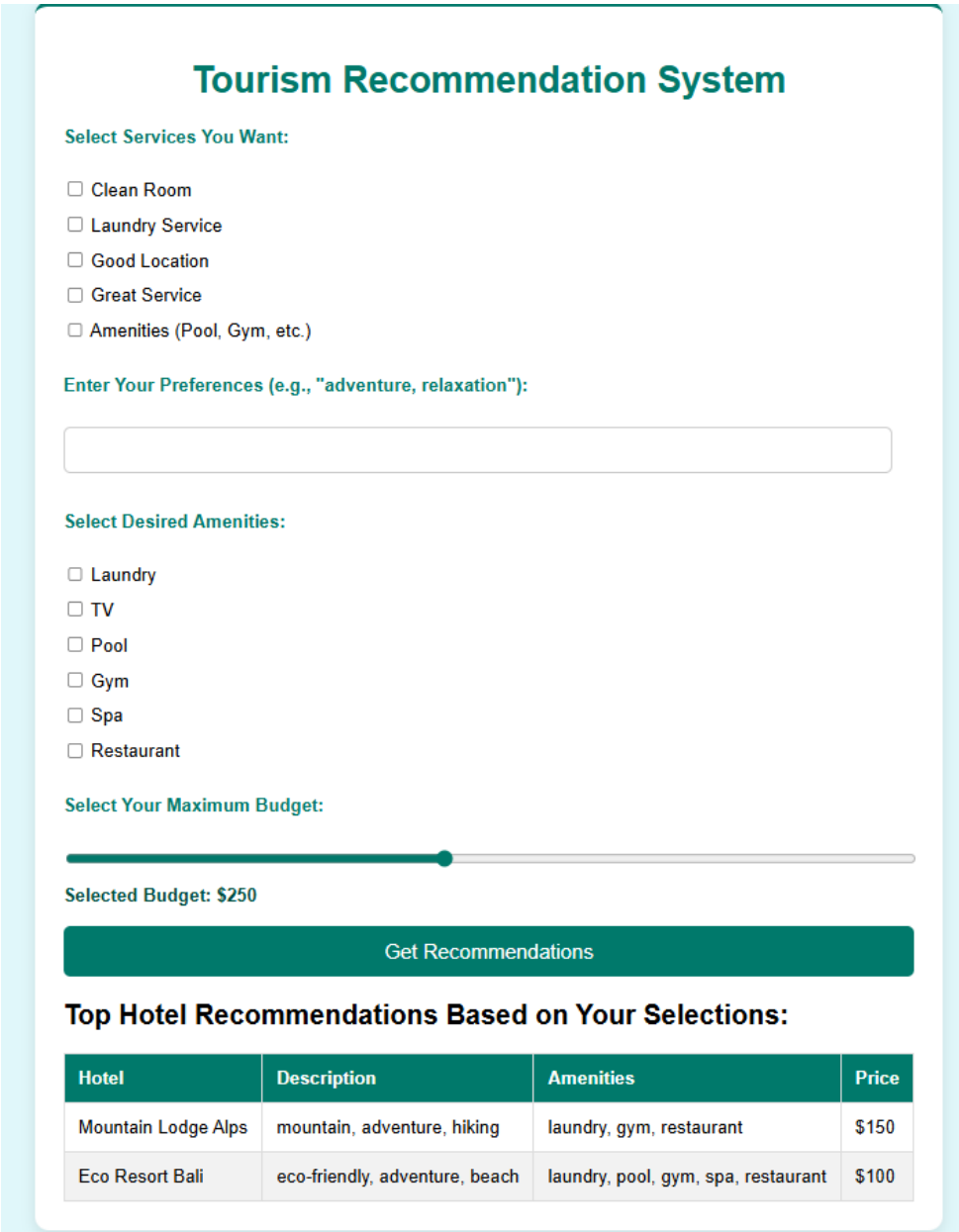


Figure 3: Tourism recommendation System user interface

3 Technology used

For the system described, the following technologies and tools are used:

3.1 Web Scraping Tools

The Instant Data Scraper extension is used for scraping Google accommodation data, including hotel descriptions, amenities, prices, and user reviews.

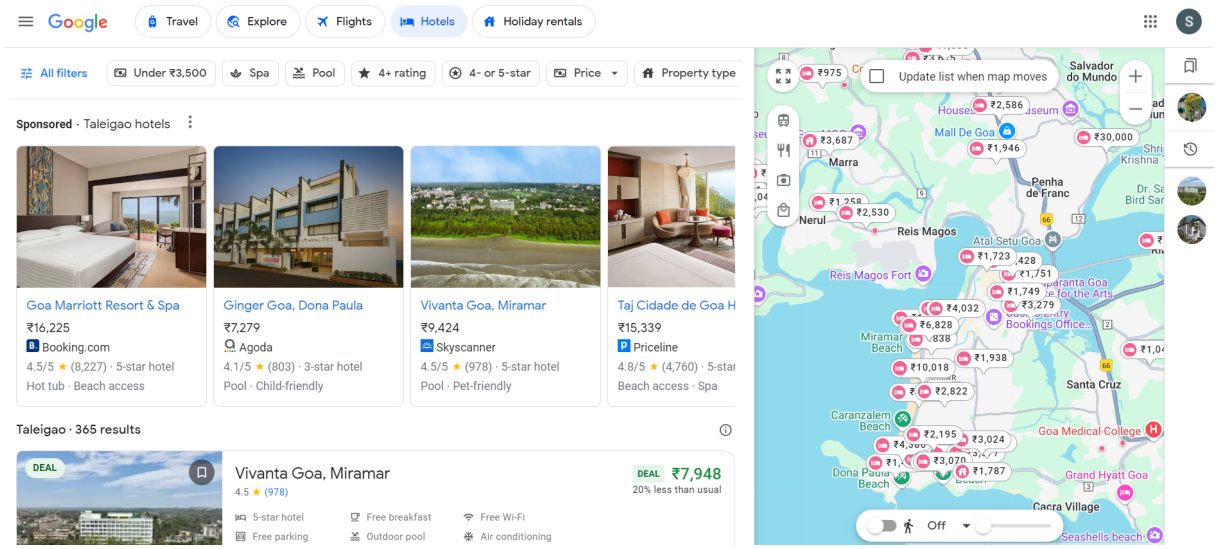
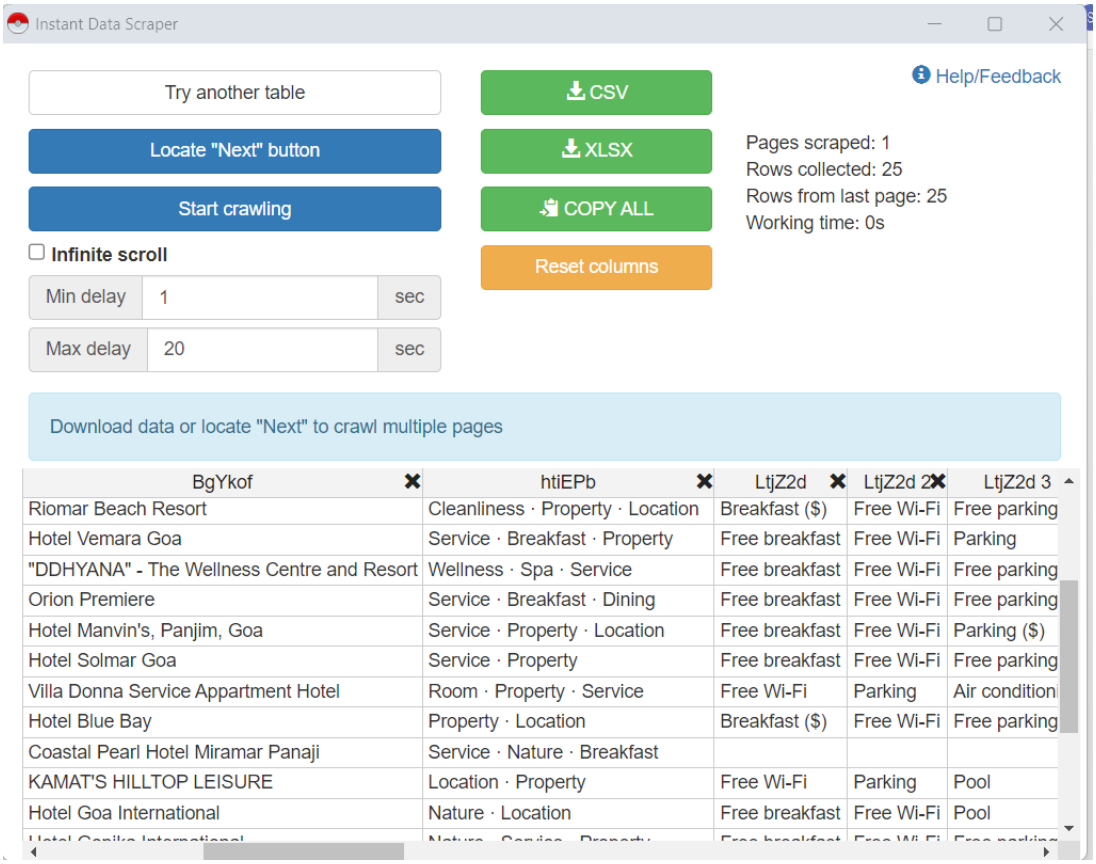


Figure 4: Hotels to recommend from Google Accommodations



BgYkof	htiEPb	LtjZ2d	LtjZ2d 2	LtjZ2d 3
Riomar Beach Resort	Cleanliness · Property · Location	Breakfast ($)	Free Wi-Fi	Free parking
Hotel Vemara Goa	Service · Breakfast · Property	Free breakfast	Free Wi-Fi	Parking
"DDHYANA" - The Wellness Centre and Resort	Wellness · Spa · Service	Free breakfast	Free Wi-Fi	Free parking
Orion Premiere	Service · Breakfast · Dining	Free breakfast	Free Wi-Fi	Free parking
Hotel Manvin's, Panjim, Goa	Service · Property · Location	Free breakfast	Free Wi-Fi	Parking ($)
Hotel Solmar Goa	Service · Property	Free breakfast	Free Wi-Fi	Free parking
Villa Donna Service Appartment Hotel	Room · Property · Service	Free Wi-Fi	Parking	Air condition
Hotel Blue Bay	Property · Location	Breakfast ($)	Free Wi-Fi	Free parking
Coastal Pearl Hotel Miramar Panaji	Service · Nature · Breakfast			
KAMAT'S HILLTOP LEISURE	Location · Property	Free Wi-Fi	Parking	Pool
Hotel Goa International	Nature · Location	Free breakfast	Free Wi-Fi	Pool

Figure 5: Instant Data Scraper used for web scraping Google Accommodation data

3.2 Data Storage

The scraped data is organized and stored in a CSV file, structured with columns such as Hotel, Description, Review, ReviewDate, Amenities, and Price for easy manipulation and analysis.

3.3 Sentiment Analysis library

TextBlob is a python library used for performing sentiment analysis. It computes sentiment scores that reflect the emotional tone of user reviews, helping determine overall user satisfaction.

3.4 Backend

A Python web framework Flask is used to develop the backend of the web application, handling user requests and interacting with the recommendation system.

These technologies and tools collectively support the functionality of the recommendation system, from data collection and sentiment analysis to user interface design and recommendation generation.

List of Figures

4 Conclusion

This report introduced a comprehensive tourism recommendation system designed to provide personalized hotel suggestions by analyzing user preferences and contextual factors. The system utilizes a combination of sentiment analysis and contextual information to enhance recommendation accuracy. User reviews are processed to extract key preferences and sentiment scores, which are then used to match with hotel features and amenities.

The effectiveness of the system was evaluated using a dataset of Google accommodation data, which includes hotel descriptions, amenities, prices, and user reviews. The system was able to successfully filter hotels based on user-defined preferences, budget constraints, and sentiment scores, providing tailored recommendations that reflect positive experiences from other travelers.

The integration of recency into the recommendation process ensures that suggestions are based on the most current user feedback, enhancing the relevance of the recommendations. The user interface, developed with Flask, Python, CSS, HTML, offers a seamless and interactive experience for users to input their preferences and view recommendations.

While the system demonstrated strong performance in delivering relevant and timely hotel recommendations, there is potential for further improvement. Future enhancements could include incorporating additional contextual factors, such as local events or traffic conditions, to provide even more precise recommendations. Additionally, expanding the system to handle group travel scenarios could further increase its utility and user satisfaction. Overall, the proposed system represents a significant advancement in tourism recommendation technology, effectively combining sentiment analysis and contextual awareness to deliver high-quality recommendations.

References

Abbasi-Moud, Z., Vahdat-Nejad, H., & Mansoor, W. (2019). Detecting tourist's preferences by sentiment analysis in smart cities. *2019 IEEE Global Conference on Internet of Things (GCIoT)*, 1–4.

Abbasi-Moud, Z., Vahdat-Nejad, H., & Sadri, J. (2021). Tourism recommendation system based on semantic clustering and sentiment analysis. *Expert Systems with Applications*, *167*, 114324. https://doi.org/https://doi.org/10.1016/j.eswa.2020.114324

Bui, H. (2020, March). Integrating sentiment analysis in recommender systems. https://doi.org/10.1007/978-3-030-43412-0_8

Jardim, S., & Mora, C. (2022). Customer reviews sentiment-based analysis and clustering for market-oriented tourism services and products development or positioning [International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021]. *Procedia*

Computer Science, *196*, 199–206. https://doi.org/https://doi.org/10.1016/j.procs.2021.12.006

Ray, B., Garain, A., & Sarkar, R. (2021). An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. *Applied Soft Computing*, *98*, 106935. https://doi.org/https://doi.org/10.1016/j.asoc.2020.106935