Travel.io : A Mixed Approach To Hotel Recommendation Systems

Meenal Gurbaxani
meenal19434@iiitd.ac.in

Nittin Yadav
nittin20093@iiitd.ac.in

Dishant Yadav
dishant20057@iiitd.ac.in

Raman Yadav
raman20111@iiitd.ac.in

Harsh
harsh20061@iiitd.ac.in

Daevaang Khairwal
daevaang20369@iiitd.ac.in

Abstract

Our innovative hotel recommendation system aims to provide personalized hotel recommendations based on user preferences. Unlike traditional approaches that rely solely on amenities and reviews, our system combines the power of machine learning algorithms with similar image detection to filter and rank hotels based on user-selected images. By incorporating parameters such as price, location, amenities, and reviews and allowing users to select images that align with their aesthetic preferences, our approach aims to offer a more tailored and visually appealing hotel recommendation experience. We understand that travellers have unique tastes and preferences when it comes to the ambience, vibe, and aesthetics of a hotel, and our system strives to cater to these individual preferences.

Code can be found here https://github.com/nittin20093/Group38_IR_project

1. Introduction

The traditional hotel booking process often lacks personalization, with recommendations based on generic factors such as relevance, paid promotions, or reviews of other customers. A novel approach has been proposed to personalize hotel recommendations based on user preferences to address this limitation. This approach integrates a mixture of two approaches: filtering hotels based on parameters such as amenities, budget, location, and reviews and incorporating a similar image detection model to filter hotels based on user-selected images.

The proposed system begins by allowing users to specify their preferences, such as price, location, and amenities. Hotels from a dataset are then filtered based on these parameters, narrowing down the options for the user. Next, the filtered hotels are ranked using a machine-learning model that takes into account user reviews and amenities as parameters. The top-ranked hotels are then presented to the user, along with images of the rooms and hotel ambience.

A unique aspect of this approach is the integration of a similar image detection model. Users are prompted to select multiple images based on their interests, which allows them to visualize the hotel's vibe, ambience, and aesthetics. These user-selected images are then used to filter the hotels further using the image detection model, which identifies similar images from the dataset. The filtered hotels are then sorted based on their rank calculated in the previous step, and the top images are presented to the user again for further selection.

This iterative process continues until the user has finalized their hotel choice or the dataset is reduced to a manageable number of images. The goal is to provide users with a personalized hotel recommendation experience that takes into consideration not only traditional parameters such as price and amenities but also the visual appeal and ambience of the hotel as captured in the user-selected images.

However, challenges such as obtaining a diverse and representative dataset of hotel images and managing scalability need to be addressed to ensure the accuracy and efficiency of the system. Despite these challenges, the proposed approach has the potential to offer a unique and personalized hotel recommendation system that considers both objective and subjective factors, providing users with a more tailored and visually appealing hotel booking experience.

2. Problem Statement

The problem we aim to address is the lack of personalization in hotel recommendations provided by popular hotel booking websites. Many hotels offer similar amenities, but they often fall short in tailoring the room according to the specific preferences and tastes of individual users. This can result in disappointing experiences for travelers who do not find the hotel rooms suitable for their needs. The current methods of hotel recommendation often do not take into account the visual preferences of users or incorporate a comprehensive ranking system. Therefore, there is a need for a recommendation system that combines visual information, user preferences, and hotel amenities to provide personalized and refined hotel recommendations.

3. Motivation

In real-life scenarios, industrial-level hotel recommendation systems generally rely on amenities and reviews from other users to suggest top hotels to a user. However, our approach combines two different methods - a ranking system based on amenities and reviews, along with a similar image detection model. We believe that by incorporating visual information and personal preferences of users, we can create a more refined and personalized hotel recommendation system. Our motivation is to provide users with a unique and customized experience in choosing hotels that align with their preferences, leading to more satisfying travel experiences.

4. Related Works

With the aim of providing recommendations to customers, Collaborative Filtering has been investigated across a variety of industries, including social media (Jiang et al., 2020), restaurants (Liu et al., 2013), and travel (Zheng, Burke, & Mobasher, 2012). The evaluation approaches used include Pearson Correlation Coefficient, Spearman Rank Correlation Coefficient, Cosine Similarity, and Mean-Square Difference in order to identify similar items or users (Pappas Popescu-Belis, 2015). In the hospitality sector, collaborative filtering is a well-liked method for creating recommendation systems, notably for hotel recommendations.

4.1. Hotel Recommendation Systems

[1] Within this paper, the context-aware hotel recommendation (CAPH) approach is considered to recommend hotels based on their features and the type of traveller who will be using them. The system's accuracy is obtained and compared to the traditional user and item-based collaborative filtering models.

[2]This paper proposes a hybrid recommendation model based on deep emotion analysis and multi-source view fusion. This model creates individualised recommendations using user-post interaction evaluations, implicit feedback, and auxiliary data. The DMHR algorithm proposed within this paper performs significantly better at score prediction and suggestion.

[3]This paper takes traditional user-based collaborative filtering and decomposes it into three context-sensitive components, and proposes a hybrid contextual approach. It can be seen that choosing an appropriate relaxation of the contextual constraints for each component of an algorithm outperforms the strict application of the context.

[4]This paper proposed a hotel recommendation system that combines collaborative filtering and the RankBoost algorithm. The system's ability to accurately recommend hotels to users was tested using reviews of hotels. The article shows the usefulness of merging various recommendation algorithms and the possibility of personalised recommendation systems in the hotel sector.

4.2. Sentimental Analysis using VADER

[1] This paper introduced VADER (Valence Aware Dictionary for sEntiment Reasoning), a rule-based sentiment analysis model that analyses social media material using a sentiment lexicon with over 9,000 lexical elements. Researchers and practitioners of natural language processing (NLP) find VADER a popular tool for sentiment analysis across various fields and domains. Multiple models like NB, ME, SVM-C and SVM-R were used .

[2]This research used VADER to analyse sentiment in restaurant reviews. According to the study, VADER had an accuracy rate of over 85% when categorising reviews as good, negative, or neutral. The study demonstrates how VADER can be used to analyse the sentiment of consumer feedback .

4.3. Image Similarity Detection

[1] This paper used TensorFlow 2.0 to create a model for image similarity detection. The model, built on a convolutional neural network, detected similar photos with high accuracy. In particular, applications where picture similarity detection is crucial, are highlighted in the research as examples of where deep learning techniques might be used for computer vision tasks.

[2] This research suggests a subject model and sentiment analysis-based visual recommendation system for peer-to-peer accommodation. To offer comparable accommodations, the study blends textual information from internet evaluations with visual elements taken from photographs. The study highlights the value of combining several recommendation techniques and the potential of visual characteristics in the lodging sector.

Overall, the use of image similarity in hotel recommendation systems shows promising results in improving recommendation accuracy and user satisfaction. Research on collaborative filtering-based hotel recommendation systems shows that this method is beneficial for providing individualised recommendations in the hospitality sector. Research in this area will likely benefit from deep learning techniques, hybrid recommendation systems, and matrix factorisation, as they are all intriguing directions for this field's future study. However, there are still challenges to be addressed, such as the high computational cost of image processing and the need for large amounts of image data.

5. NOVELTY

The system has undergone further personalization through the incorporation of an additional ranking criterion based on location. Specifically, the system considers the ranking of pincodes or postal codes in alignment with customer requirements, such as tourism, and integrates this parameter into the overall ranking algorithm.

Our system allows users to input sample images, which are then analyzed using computer vision algorithms to identify visual features. Based on these visual features, the system recommends hotels that share similar aesthetics. This feature is especially beneficial for users with specific aesthetic preferences or those seeking hotels with a particular style.

6. DATA COLLECTION AND PRE-PROCESSING

6.1. Data Collection

For our dataset collection, we web scraped hotel data from websites such as Oyo and Trivago using the Selenium and Beautiful Soup (BS4) libraries. We extracted raw data and then cleaned it during pre-processing, removing unwanted amenities and merging repeated amenities with different names to improve training and results. The unformatted data was sorted and stored in CSV files.

6.2. Data Pre-Processing

During pre-processing, the extracted data was in the form of a dictionary. We cleaned it and stored it in a data frame with dedicated columns for each feature. We then extracted all the amenities of all the hotels and created dedicated columns for them. We merged amenities with similar meanings or synonym words, such as TV and television, and assigned them scores of 0 and 1 to represent their presence or absence in each hotel. This information was stored in a data frame.

Next, we calculated the frequency of occurrences for each amenity and created a sorted dictionary. From this dictionary, we selected 25 amenities that were deemed more significant. Among these 25 amenities, we chose 13 with the highest frequency and 12 from the middle range, as amenities with low frequency were considered unnecessary and time-consuming for users. We removed all other amenities from the data and cleaned it again, resulting in a final data frame with only 25 significant amenities for the recommendation system. This final data frame was then converted to a CSV file for further processing.

7. METHODOLOGY

7.1. RANKING SYSTEM

Our first step after preprocessing the data is to create a ranking system which will then be filtered based on the needs and requirements of amenities by the customer. For said ranking system, we began by creating scores for each of the hotels mentioned depending upon the basis of reviews and ratings provided. For proper analysis of the text given in the reviews, we used VADER (Valence Aware Dictionary and sEntiment Reasoner)

VADER is a rule-based sentiment analysis model that uses a sentiment lexicon to calculate the polarity of the text and was introduced by researchers Hutto and Gilbert in 2014. It analyses text, including social media writing, using a sentiment vocabulary with different sentiment intensities. 9,000 lexical qualities have now been added to VADER's vocabulary using automated and hand annotation processes. Several tests have shown that the model's performance is on par with or even better than other sentiment analysis algorithms. VADER has been used for political analysis, maintaining brand reputation, and assessing customer feedback since it is computationally efficient.

7.1.1 Base Scores according to Reviews and Ratings

Based on the text's polarity, the algorithm gives each review a sentiment score; a positive score denotes a positive feeling, and a negative score indicates a negative sentiment. Once we obtained the sentiment scores provided by the algorithm for each review, we then used the corresponding ratings as weights and multiplied them to each other. We then added the results and generated a single aggregate score for each hotel in our dataset.

	Unnamed: 0.2	details	Location	Name	Score	reviews	price	img_link	room_size	amenities	...	Bathrobes	Adapter	Slippers
406	406	{'Location': 'No 1380 1 Near Morjim Football G...	No 1380 1 Near Morjim Football Ground, Tembwad...	Tropical Wave Hostel	9.912	[{'Review_Rating': '10.0', 'Review_Heading': '...	2,498	https://q-xx.bstatic.com/xdata/images/hotel/84...	14m²	['Balcony/terrace', 'Shower', 'Free Wi-Fi', 'T...	...	0	0	0
415	415	{'Location': 'Arambol Beach, Goa, India', 'ove...	Arambol Beach, Goa, India	Goym Resort	9.814	[{'Review_Rating': '10.0', 'Review_Heading': '...	3,820	https://pix8.agoda.net/hotelImages/35712803/58...	10m²	['Balcony/terrace', 'Shower', 'Free Wi-Fi', 'T...	...	0	0	0
465	465	{'Location': "Tito's lane 2, Saunta Vaddo, Bag...	Tito's lane 2, Saunta Vaddo, Baga Near Snow Pa...	Baga Bao	9.783	[{'Review_Rating': '10.0', 'Review_Heading': '...	1,800	https://q-xx.bstatic.com/xdata/images/hotel/84...	0	['Shower and bathtub', 'Free Wi-Fi', 'Toiletri...	...	0	0	0
183	183	{'Location': 'Cola Goa Beach Resort, Agonda, G...	Cola Goa Beach Resort, Agonda, Goa, India, 403702	Cola Goa Beach Resort	9.694	[{'Review_Rating': '10.0', 'Review_Heading': '...	20,900	https://q-xx.bstatic.com/xdata/images/hotel/84...	30m²	['Balcony/terrace', 'Shower and bathtub', 'Fre...	...	1	1	1
410	410	{'Location': 'Survey No, 155/1, Arpora Baga, O...	Survey No, 155/1, Arpora Baga, Opp - Radisson ...	Caravan Baga Aqua Resort	9.694	[{'Review_Rating': '10.0', 'Review_Heading': '...	3,000	https://q-xx.bstatic.com/xdata/images/hotel/84...	30m²	['Balcony/terrace', 'Shower', 'Toiletries', 'T...	...	0	0	0

Figure 1. Top 5 hotels depending on the ranking of Base Scores

7.1.2 Personalised Scores on the Basis of Tourism

After obtaining the aggregate score based on reviews and ratings, we have created a ranking of the pin codes of the top ten tourist destinations within our specified cities. This list of ranks also helps provide a reference for tourism-specific hotel rankings by increasing the aggregate scores when a hotel belongs to the specified pin code area.

	Unnamed: 0.2	details	Location	Name	TouristScore	Score	reviews	price	img_link	room_size	...	Bathrobes	Adapter	Slippers
410	410	{'Location': 'Survey No, 155/1, Arpora Baga, O...	Survey No, 155/1, Arpora Baga, Opp - Radisson ...	Caravan Baga Aqua Resort	18.418600	9.684000	[{'Review_Rating': '10.0', 'Review_Heading': '...	3,000	https://q-xx.bstatic.com/xdata/images/hotel/84...	30m²	...	0	0	0
335	335	{'Location': 'Behind Campal Garden, Campal, Pa...	Behind Campal Garden, Campal, Panjim, Goa, Pan...	Surya Kiran Heritage Hotel	18.401400	9.200700	[{'Review_Rating': '9.6', 'Review_Heading': 'G...	6,579	https://pix8.agoda.net/hotelImages/18982190/-1...	26m²	...	0	1	1
479	479	{'Location': 'Opposite, St. Joseph School, Ca...	Opposite, St. Joseph School., Calangute, Goa, ...	Cassa-e-DiWaN	16.826248	8.855920	[{'Review_Rating': '9.2', 'Review_Heading': 'S...	3,999	https://q-xx.bstatic.com/xdata/images/hotel/84...	19m²	...	0	0	0
339	339	{'Location': 'Off Assagaon Badem Rd, Vagator, ...	Off Assagaon Badem Rd, Vagator, Goa, India, 40...	Mademoiselle Boutique Hotel and Cafe	14.926400	9.329000	[{'Review_Rating': '10.0', 'Review_Heading': '...	5,599	https://pix8.agoda.net/hotelImages/07124650/-1...	27m²	...	0	0	1
256	256	{'Location': '215/17 khobra vaddo, opposite to...	215/17 khobra vaddo, opposite to phoneshwar to...	Pajaros Blu	14.841787	7.811497	[{'Review_Rating': '10.0', 'Review_Heading': '...	4,417	https://q-xx.bstatic.com/xdata/images/hotel/84...	33m²	...	1	1	1

5 rows × 30 columns

Figure 2. Top 5 hotels depending on the ranking of Tourism Scores

7.1.3 Final Rankings and Amenities Selection

These scores obtained were stored within a column alongside our dataset and then were used to rank the hotels. Hotels with greater scores are ranked higher; thus, we obtain our ranked dataset.

Ranked Hotel Datasets

This dataset will further be reduced when the customer selects specific criteria (price, location, amenities etc.). Only the hotels fulfilling those criteria will be passed on for the image similarity step.

7.2. SIMILAR IMAGE DETECTION

For Similar image detection, We are using the Tensorflow library's "Tensorflow Hub" to generate 'image feature vectors' of the images of hotels. Then we are planning to use the "Annoy" library and image feature vectors to calculate the similarity scores and filter out hotels based on similarity scores.

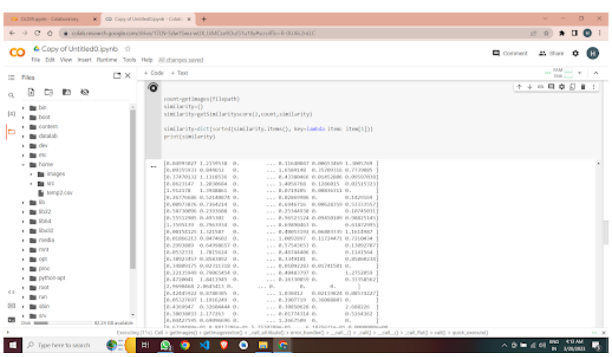

Figure 3. vector representation of image which will further will be used as similarity detection

8. CHALLENGES

While VADER sentiment analysis excels at categorizing a wide range of syntactical patterns, it has a limitation in that it solely focuses on individual words without taking into consideration their surrounding context. As a rule-based classifier, VADER can only recognize specific scenarios that are pre-programmed into its algorithm. This means that it can account for the contextual effects of words like "but," but it may overlook conjunctions such as "although" or "however."

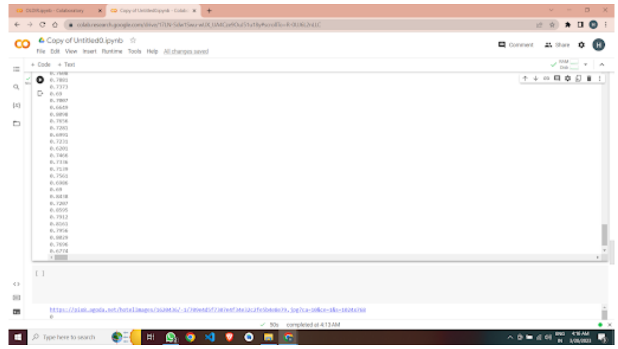

Figure 4. This is the similarity of the image as output with others in a serial manner

9. CONCLUSION AND FUTURE WORKS

Our website incorporates a parameter-based selection process that involves choosing parameters such as price, location, and specific amenities. Subsequently, the dataset is filtered, and the filtered dataset is then ranked, with the images of the best results displayed at the top. Users have the option to select some of the photos, and a similar image detection algorithm is executed on the dataset. Additional photos relevant to the earlier choices are shown repeatedly until the user is satisfied with the photo selection. The process can be terminated directly at the request of the user or when the dataset has been exhausted.

Furthermore, we propose integrating social media data from social media networks to provide recommendations based on user activity and preferences. We also aim to improve the visualization of hotel amenities and locations to assist users in making more informed decisions. Additionally, we may integrate with other travel service providers, such as flight and vehicle rental companies, to offer a more comprehensive travel planning experience. Lastly, we intend to explore incorporating voice commands and other smart device features to enable consumers to make hotel reservations and receive recommendations.

This idea can be extended to other areas that play an important role in our day-to-day lives, such as choosing a restaurant, purchasing a new home, or renting a house, among others.

10. INDIVIDUAL CONTRIBUTION

Nittin Yadav - Web Scraping Data Collection
Meenal Gurbaxani - Ranking System, Literature Review
Dishant Yadav - Data Preprocessing
Raman Yadav - Frontend UI
Harsh - Similar Image Detection
Daevaang Khairwal - Frontend UI

11. REFERENCES

[1] Hotel recommendation system based on review and context information: A ... (n.d.). Retrieved March 27, 2023, from https://core.ac.uk/download/pdf/301369618.pdf

[2] Differential context relaxation for context-aware travel ... - researchgate. (n.d.). Retrieved March 27, 2023, from https://www.researchgate.net/publication/230773488_Differential_Context_Relaxation_for_Context-Aware_Travel_Recommendation

[3] Jiang, L., Liu, L., Yao, J., Shi, L. (2020, October 7). A hybrid recommendation model in social media based on deep emotion analysis and multi-source view Fusion - Journal of Cloud Computing. SpringerOpen. Retrieved March 27, 2023, from https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-020-00199-2

[4] A hotel recommendation system based on collaborative filtering and ... (n.d.). Retrieved March 27, 2023, from https://www.researchgate.net/publication/238516069_A_Hotel_Recommendation_System_Based_on_Collaborative_Filtering_and_Rankboost_Algorithm

[5] Hutto, C., Gilbert, E. (n.d.). Vader: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media. Retrieved March 27, 2023, from https://doi.org/10.1609/icwsm.v8i1.14550

[6] Beri, A. (2020, May 27). Sentimental analysis using vader. Medium. Retrieved March 27, 2023, from https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664

[7] Isbilen, E. (2019, December 29). Image similarity detection in action with tensorflow 2.0. Medium. Retrieved March 27, 2023, from https://towardsdatascience.com/image-similarity-detection-in-action-with-tensorflow-2-0-b8d9a78b2509

[8] Li, D., Yin, H., Wang, C., Song, S., Li, K., Li, C. (2022). Visual recommendation for peer-to-peer accommodation with online reviews based on sentiment analysis and topic models. Journal of visualization. Retrieved March 27, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9127497/

[9] Journal, I. R. J. E. T. (2017, June 20). Hotel recommendation based on User Preference Analysis. 2015 31st IEEE International Conference on Data Engineering Workshops. Retrieved February 21, 2023, from https://www.academia.edu/33567241/Hotel_Recommendation_based_on_User_Preference_Analysis

[10] Computing similarity on images using machine learning and cosine similarity. Computing Similarity on Images using Machine Learning and Cosine Similarity. (2020, July 10). Retrieved March 13, 2023, from https://soumilshah1995.blogspot.com/2020/07/computing-similarity-on-images-using.html

[11] Author links open overlay panelJianshan Sun a, a, b, c, websites, A. S. media, Xia, R., Wang, G., Liu, H., Liu, J. G., Chen, L., Abbasi, A., Aciar, S., Adomavicius, G., Aspden, P., Baccianella, S., Balahur, A., Banker, C., Bell, R., Bennett, J., ... Li, Y. (2014, October 27). Mining affective text to improve social media item recommendation. Information Processing Management. Retrieved March 13, 2023, from https://www.sciencedirect.com/science/article/abs/pii/S0306457314000879

[12] Popescu-belis, A. (2015, August 3). Combining content with user preferences for non-fiction Multimedia recommendation: A study on ted lectures. Academia.edu. Retrieved March 13, 2023, from https://www.academia.edu/14599858/Combining_content_with_user_preferences_for_non_fiction_multimedia_recommendation_a_study_on_TED_lectures

[13] A hotel recommendation system based on collaborative filtering and ... (n.d.). Retrieved March 13, 2023, from https://ieeexplore.ieee.org/document/5474286