
SMART TRAVEL ASSISTANT WITH ITINERARY PLANNER USING HYBRID MACHINE LEARNING APPROACH

Mrs. S. Sangeetha Mariammal^{*1}, S B Akshaya^{*2}, M Priyanga^{*3},
S Saran Kumar^{*4}, P Prakash^{*5}

^{*1}Assistant Professor, Department Of Computer Science And Engineering, Coimbatore
Institute Of Technology, India.

^{*2,3,4,5}Student, Department Of Computer Science And Engineering, Coimbatore Institute
Of Technology, India.

ABSTRACT

Tourism is an important industry and a popular recreational activity done by millions around the world. Selecting a tourist destination from the available information is one of the most difficult tasks for tourists while planning travel, before and during their trip. Also, another important task for visitors is to plan and organize travel routes that include many points of interest (POIs). In this proposed system, the unique preferences of each user are taken into account and the system aims to guide the visitor in arranging the route according to his/her areas of interest. The database is collected from tourism review websites such as TripAdvisor & Holidify using web scraping and the YELP database is used for reference.

This approach gathers URLs of various tourist attractions from TripAdvisor and Holidify and in stages, information about tourist attractions and reviews will be pulled from the URLs collected. K-Means clustering and KNN methods are applied to cluster nearby attractions and hotels, Logistic Regression to perform Sentimental Analysis on users' comments, and Transfer Learning using the VGG 16 module to locate the place where the image provided by the user was captured, a Hybrid ML model will be constructed. The generated model will be trained and tuned based on the features derived from the data collected. As a result, a best model will emerge that can create numerous suggestions and itineraries for consumers.

Keywords: Hybrid Machine Learning, Recommendation, Travel Itinerary, Sentiment Analysis, Logistic Regression, K-Means, K-Nn, Transfer Learning, Vgg 16, 2-Opt Algorithm.

I. INTRODUCTION

The travel industry is continuously on the cutting edge of technological advancements. Tourism contributed \$194 billion to India's GDP in 2019, according to the World Travel and Tourism Council (WTTC). In 2020, the Indian tourism industry will employ 31.8 million people, accounting for 7.3 percent of the country's overall workforce. It is estimated to employ almost 53 million people by 2029. As a result, developing a better web-based product that provides information about various tourist sites, means of transportation, and lodging based on customers' preferences can help the Indian economy flourish while also increasing customer satisfaction. The aim of this project is to develop an one-stop website for tourists to make travel plans and to schedule customized tour itineraries which has not been advanced so far as a business model in Indian tourism sector.

The two main goals of this paper are to create a smart travel system that uses machine learning to categorise and recommend various tourist destinations to users based on their preferences, as well as to provide a customised itinerary that includes possible places to stay, inns, and attractions, and to dynamically update the recommendations based on user feedback.

II. METHODOLOGY

SYSTEM ARCHITECTURE :

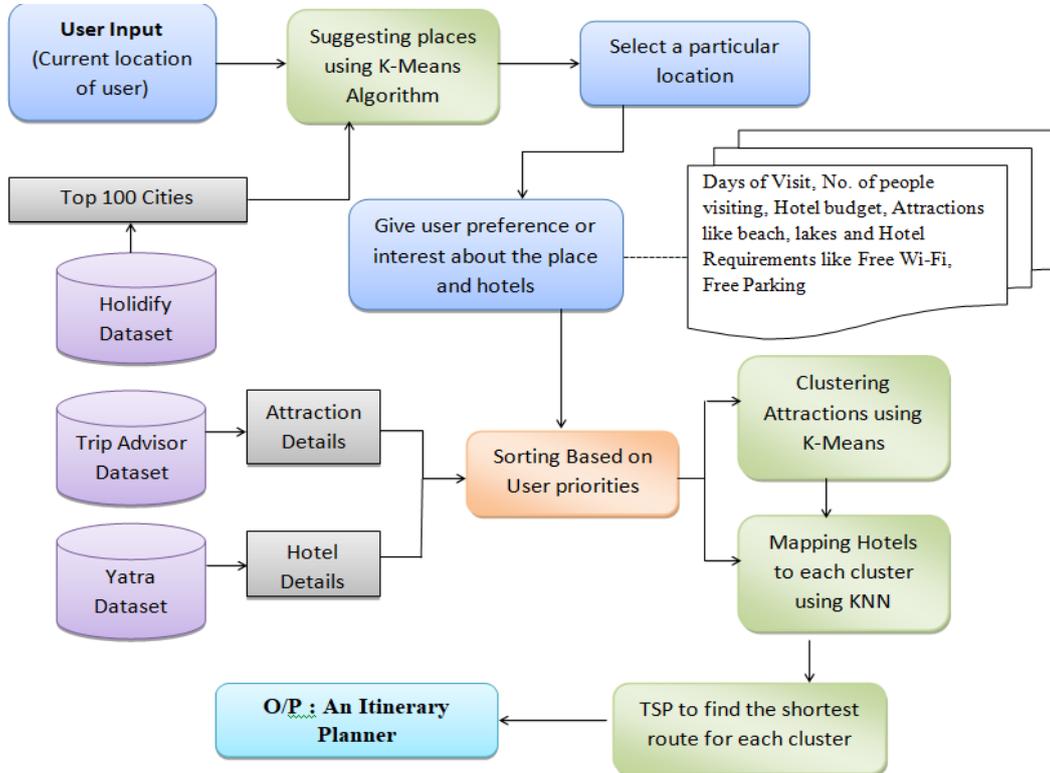


Figure 1: System Architecture

DATA COLLECTION :

Dataset is not obtained from any of the sources like Kaggle because there is no availability of information related to tourist attractions or hotels that are only based in India. Dataset has been prepared by scraping information from some of the tourism related websites like Holidayfy, TripAdvisor and Yatra. Using Holidayfy, top 100 tourist destinations in India are collected. TripAdvisor is used to get details of tourist attractions and the attributes included are name of the tourist attraction, the city in which it is located, the type of tourism it belongs to, its rating and the total number of reviewers.

For collecting data related to hotels, the website Yatra is used and the details collected from this website are the name of the hotel, the city where it is located, type of hotel/rooms available, its price, its rating and total number of reviewers. Apart from this latitude and longitude of each attraction and hotel are collected from Google and are saved in the respective datasets. Also a derived parameter namely, "total_score" is obtained by multiplying the rating and the total number of reviewers to know its actual weightage.

RECOMMENDATION ENGINE :

Recommendation engines are constructed using a hybrid machine learning approach for suggesting tourist destinations that are nearer to the user's current location and also for recommending travel itineraries for the selected destination based on user preferences.

Various machine learning algorithms are implemented in the proposed system in a hybrid manner and are explained in detail below :

K - NEAREST NEIGHBOR ALGORITHM :

K-Nearest Neighbor is a Supervised Learning-based Machine Learning algorithm that is one of the most basic. The K-NN method assumes that the new case/data and existing cases are similar and places the new case in the category that is most similar to the existing categories. It saves all of the available data and classifies a new data point based on its resemblance to the existing data. This means that new data can be quickly sorted into a suitable category using the K-NN method. It can be used for both regression and classification, but it is more commonly employed for classification problems.

K - MEANS ALGORITHM :

This unsupervised Machine Learning algorithm called K-Means Clustering divides the unlabeled dataset into various clusters. K specifies the number of predefined clusters that must be created during the process; for example, if K=2, two clusters will be created, and if K=3, three clusters will be created, and so on. It provides a simple approach to cluster data into multiple groups and a quick way to find the categories of groups in an unlabeled dataset without any training. Each cluster is paired with a centroid in this centroid-based technique. The K-Means algorithm's main goal is to reduce the sum of distances between points and their corresponding cluster centroid.

OPT ALGORITHM :

The 2-Opt algorithm belongs to the local search family. This algorithm begins with an initial solution and iteratively seeks out chances for improvement in the vicinity of that solution. This initial answer can be of any type as long as it is practicable. The 2-opt algorithm works by taking two arcs from the route, connecting them together, and calculating a new journey distance. The current route is updated if this alteration resulted in a shorter total trip distance. The algorithm builds on the improved route by repeating the stages. This method is repeated until no further improvements are discovered or a predetermined number of iterations have been completed.

It has a unique swapping mechanism that functions as a heuristic. The 2-opt method's major goal is to eliminate path crossing in each city's neighborhood. The 2-opt strategy is simple to implement and execute. This approach, for example, can solve TSP with 120 cities in less than 5 seconds on an Intel Core i7 processor. "Solved" here refers to the algorithm arriving at a good-enough but sub-optimal result. 2-opt removes two edges and adds two new ones (assuming the cost matrix is symmetric), allowing the cost function to be simplified to only consider the changing edges. This is substantially faster than enumerating over the entire path for huge arrays.

LOGISTIC REGRESSION :

A strong supervised machine learning approach is logistic regression. When the data in question provides binary output, such as whether it belongs to one of two classes or is either a 0 or 1, it is one of the most frequent binary classification approaches. It's an S-shaped curve that can convert any real-valued number to a number between 0 and 1. When utilizing logistic regression, a threshold is frequently set, indicating at what point the example will be classified as one of two classes.

Logistic regression models the data using the Logistic or Sigmoid function.

$$y = 1 / (1 + e^{-x})$$

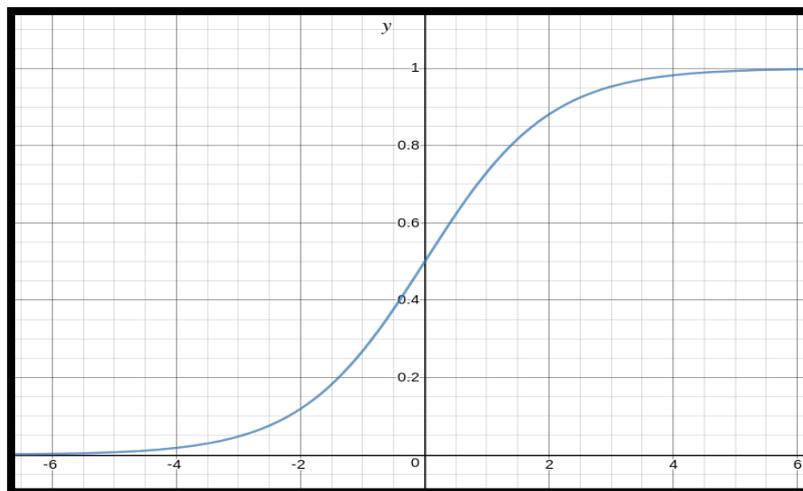


Figure 2: Graph for Sigmoid Function

Inference from the above graph :

- As x -> infinity, y tends to 1.
- y goes to 0 as x approaches (-infinity)

- y is always confined between 0 and 1.

TRANSFER LEARNING :

Transfer learning is the practice of using a previously learned model to solve a new problem. In transfer learning, a machine uses knowledge from a previous assignment to improve prediction about a new task. During transfer learning, the information of an already trained machine learning model is transferred to a separate but closely related task. We might utilize the model's training experience to detect other things such as sunglasses if we trained a simple classifier to predict whether an image contains a backpack. Transfer learning is commonly used in computer vision and natural language processing tasks like sentiment analysis due to the enormous amount of CPU resources required. Transfer learning has several advantages, the most prominent of which are shorter training times, better neural network performance (in most cases), and the lack of a huge amount of data.

➤ **VGG-16 using Transfer Learning Approach :**

VGG-16 is a 16-layer deep convolutional neural network. It's a 92.7 percent accurate object identification and classification system that can categorize 1000 photos into 1000 different categories. It is a common picture classification technique that is simple to employ with transfer learning. We can use the ImageNet database to load a pre-trained version of the network that has been trained on over a million photos. The network has been pre-trained to classify photos into 1000 object categories, including keyboards, mice, pencils, and a variety of animals. As a result, the network has learned a variety of rich feature representations for a variety of images. On the ImageNet dataset, VGG16 was shown to be the highest performing model out of all the setups.

III. MODELING AND ANALYSIS

The proposed system attempts to address the gap left by existing systems by creating a website with a high user-interface that can generate both tour recommendations and itinerary planners for the specified tourist location in India. Many people nowadays choose to book vacations using online platforms rather than through travel agents. This method was conceived and constructed utilizing a hybrid machine learning approach after seeing the economic benefit or contribution that an online tourism website may make.

When a user chooses a travel destination, an itinerary will be generated based on user preferences such as number of vacation days, hotel budget etc.. Also an Image Finder module is implemented which will let users to find the name or location of the image that has been uploaded. Let's see each of the modules in detail.

HOME PAGE :

The top element of the Home Page module contains the website's name and a link to the Image Finder Module. In addition, there is an input field on the main screen where the user must type in his or her present location. There will be several photos, each depicting a different type of tourism attraction, such as beaches or mountains. When a user types in a location and clicks an image, the user is taken to the place recommendation module, which displays all nearby places that are also part of the type of attraction selected.

PLACE RECOMMENDATION MODULE:

This module shows the output of the preceding module, namely the locations filtered from the home page after applying the K-Means method to feed the inputs. These destinations are displayed as photos and include information such as the name of the location, the various modes of transportation accessible to get there, and the distance from the user's current location. Each destination also includes a "Explore" button that takes you to the User Preference Module when you click it.

USER PREFERENCE MODULE :

In this module, a user's tour preferences for a specific destination are acquired. This page contains various input fields such as the number of days of visit, the number of people visiting, the hotel budget, a checkbox-based input where a user can select multiple types of attractions such as lakes, mountains, religious sites, theme parks, and other attractions that he or she wishes to visit in the selected destination, and another checkbox with various hotel requirements such as free parking, free wifi, couple friendliness, and so on from which the user can select the type of hotel he or she wishes to stay in. When all of the following inputs are completed and submitted, it is redirected to the itinerary recommendation module, which receives all of the information.

ITINERARY RECOMMENDATION MODULE :

Multiple points of interest (POIs) in the specified destination will be filtered, sorted, and grouped based on user priorities using the K-Means Algorithm, and itineraries will be generated after all user preferences have been entered. If the number of days to visit equals n, the travel arrangements for days 1 through n will be provided. A cluster of POIs will be assigned to each travel plan, and a hotel that is closer to the cluster will be mapped to it using the KNN method. And, with each journey plan, a map will be displayed using the Python Folium package, which will show all of the POIs with a pin and a path connecting them.

IMAGE FINDER MODULE :

When the Image Finder link is clicked from the header of the home page, it gets redirected to the Image Finder module where it has an input field which accepts an image file. When an image is uploaded in the page, the name of the place which is captured in the image and its location will be displayed along with similar photos which are already stored in the system. This photo search engine is executed using VGG-16 and this comes under transfer learning.

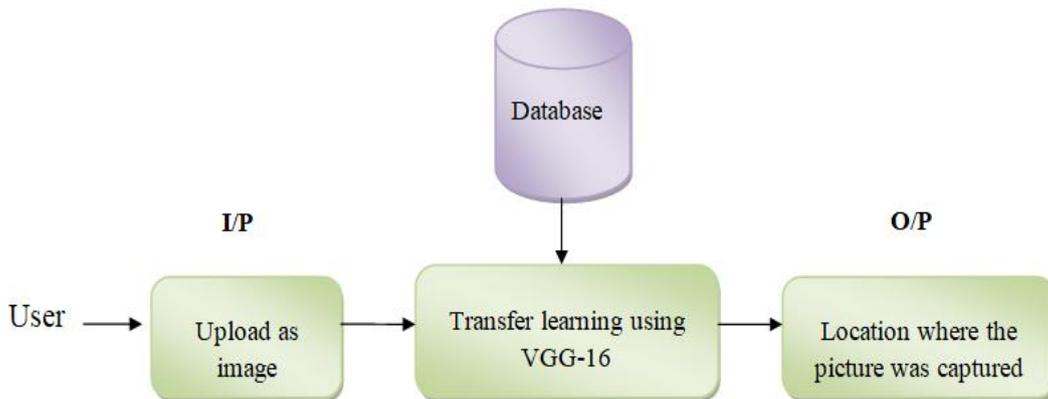


Figure 3: Workflow of Image Finder Module

FEEDBACK AND REVIEW MODULE :

When a user click the feedback button in the itinerary module, he or she will be taken to the Feedback and Review module. This module assists in gathering genuine user experience regarding their locations in order to improve itinerary advice utilizing sentimental analysis and the Logistic Regression method. When a user provides favorable feedback, tourist sites that are similar to the selected destination in terms of climatic and geographic circumstances are recommended; nevertheless, if a user provides bad feedback, the user is routed to the home page.

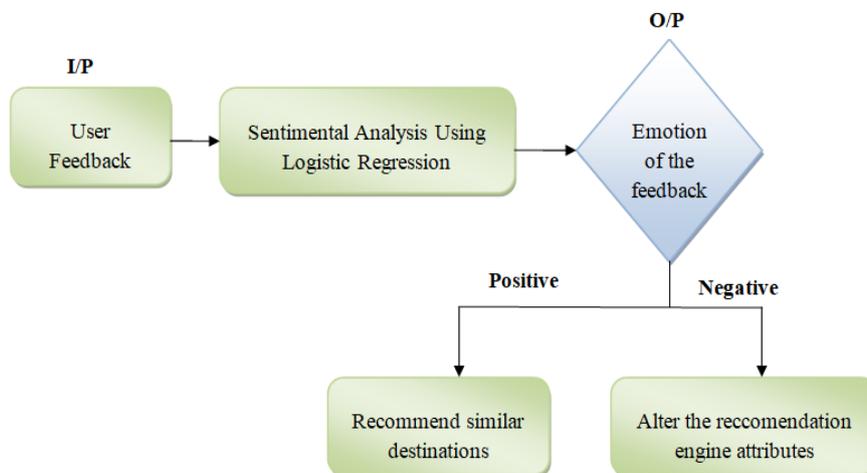


Figure 4: Workflow of Feedback and Review Module

IV. RESULTS AND DISCUSSION

In the previous chapter, all the modules that are implemented in the system have been explained in detail. The result of each module will be explained and displayed in this chapter.

✓ If the current city field on the home page is set to 'Coimbatore,' and the image of a beach is selected, all the places that are closer to Coimbatore and have a beach as their main tourist attraction will be displayed in this module.

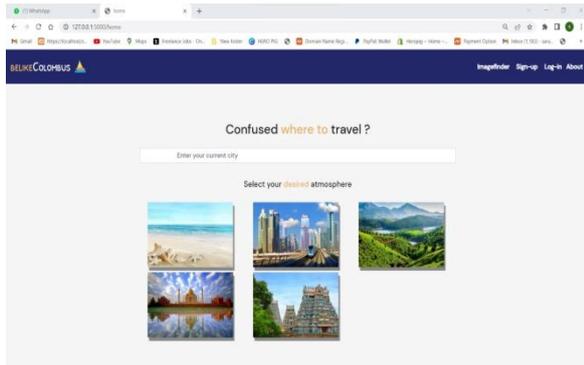


Figure 5: Home Page

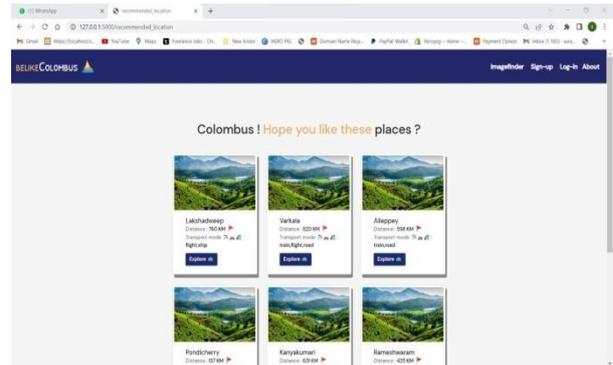


Figure 6: Place Recommendation Module

✓ In the Itinerary Recommendation module, an itinerary is generated based on the user's preferences for tourist attractions, hotels, and a few other attributes. Each day's itinerary includes a few points of interest and a hotel, and these locations are displayed in a map illustration with a path encircling all of the POIs. There will also be a few other places mentioned as additional attractions that people can visit if they have some spare time. This module also includes a Google Translator, allowing the user to experience the web page in the language that he or she is most comfortable with.

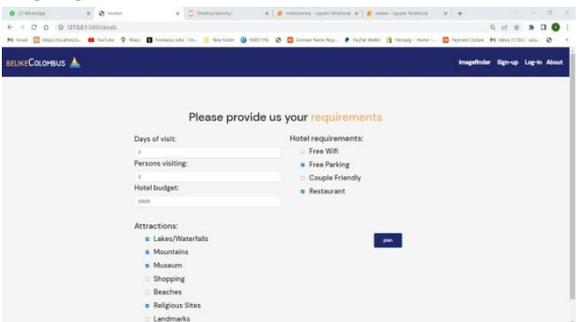


Figure 7: User Preference Module

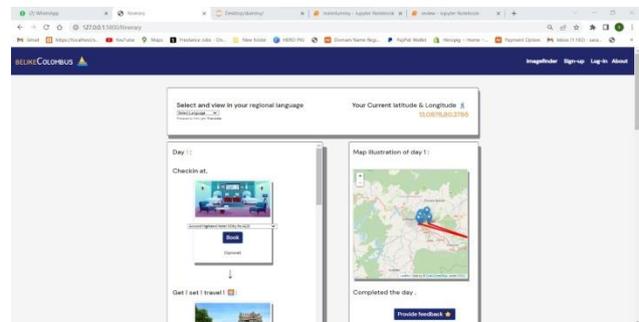


Figure 8: Itinerary Recommendation Module

✓ When an image file is uploaded in the Image Finder module's input field, the image, along with similar photos identified in the dataset, its name, and location, will be presented in the web page as a result.

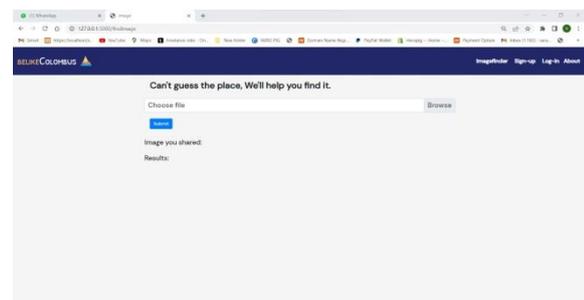


Figure 9: Image Finder Module

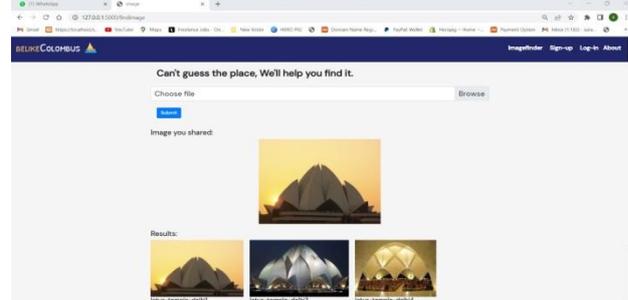


Figure 10: Result of Image Finder Module

✓ In the feedback module, when a user gives a positive review, he / she will be taken to another page where tourist places similar to the destination selected by the user in terms of climatic and geographic conditions will be suggested. Otherwise, when the user submits a negative review, the module will let the user go back to the home page.

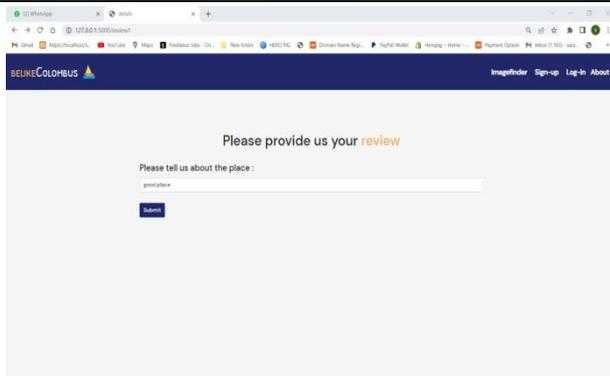


Figure 11: Feedback Module

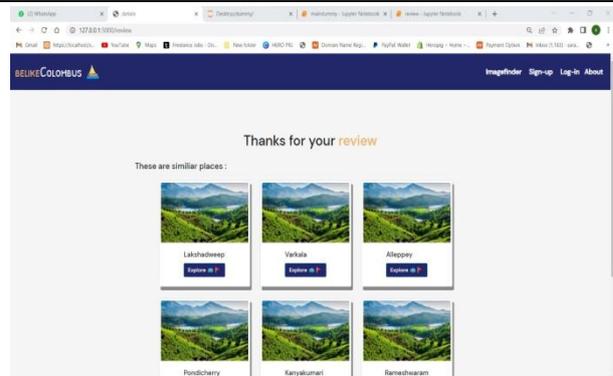


Figure 12: Output of feedback module for a Positive review

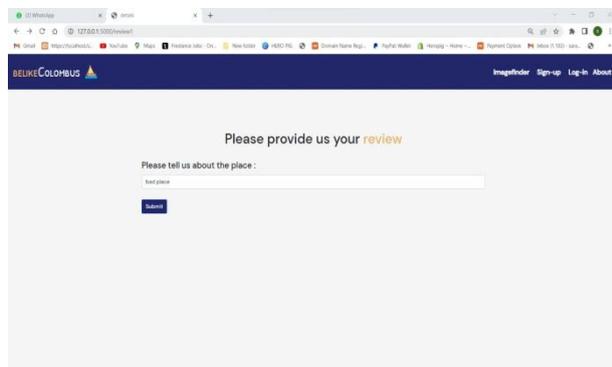


Figure 13: Feedback Module

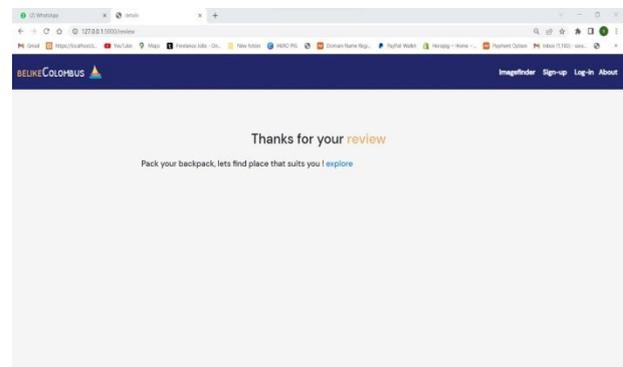


Figure 14: Output of feedback module for a negative review

V. CONCLUSION

Everything has grown up with technology in recent years. Despite this, contacting a travel agency to arrange the trip or following a fixed itinerary offered by an agency through some websites is part of the actual planning process while looking for vacations online. The proposed system uses a hybrid machine learning technique to personalize this plan as well as trip recommendations based on the preferences of the users.

Although recommendation algorithms have demonstrated their effectiveness, further development of a proposed system from a commercial standpoint can be considered future work. Another potential future direction could be to provide a module/space for users to submit their trip blogs and vlogs. This will result in increased website utilization, turning it into a commercial product.

VI. REFERENCES

- [1] Achin Jain, Vanita Jain and Nidhi Kapoor, (2016) "A Literature Survey On Recommendation System Based On Sentimental Analysis", Vol.3, No.1.
- [2] Brindha Devi V, Meenaloshini M, Sriram G and Subhashini R, (2021) "Intelligent Tourist System For The 21st Century".
- [3] Catherine Almira, Nur Ulfa Maulidevi, (2019) "Travel Itinerary Recommendation for Real World Point of Interests Using Iterated Local Search".
- [4] Omamah Alnogaithan, Sumaiah Algazlan, Aljoharah Aljuraiban and Amal A. Shargabi, (2019) "Tourism Recommendation System Based on User Reviews".
- [5] Srikanth Tammina, (2019) "Transfer Learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images", Volume 9, Issue 10, October
- [6] Swati Gajakosh, Bhagyashree Patil, Jayashri Gulve, Ravina Patil and Pooja Patil, (2019) "Travel Route Recommendation Using K-NN Algorithm", Volume 4 Issue 1.
- [7] Wai Chong Chia, Lee Seng Yeong, Fennie Jia Xian Lee, Sue Inn Ch'ng, (2016) "Trip Planning Route Optimization with Operating Hour and Duration of Stay Constraints".