A Project Report on

Personalized Itinerary Recommendation System

Submitted in complete fulfillment of the requirements for the award
of the degree of

Bachelor of Engineering

in

Computer Engineering

by

Nakulesh Jayakrishnan(17202012)
Gauri Deshpande(16102046)
Akanksha Koshti(16102011)

Under the Guidance of

Prof. Sachin Takmare

Department of Computer Engineering
A.P. Shah Institute of Technology
G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615
UNIVERSITY OF MUMBAI

Academic Year 2019-2020

Approval Sheet

This Project Report entitled ***"Personalized itinerary recommendation system"*** Submitted by ***Nakulesh Jayakrishnan(17202012), Gauri Deshpande(16202046), Akanksha koshti(16102011)*** is approved for the complete fulfillment of the requirement for the award of the degree of ***Bachelor of Engineering*** in ***Computer Engineering*** from ***University of Mumbai***.

(Prof. Sachin Takmare)
Guide

(Prof. Sachin Malave)
Head Department of Computer

Place:A.P.Shah Institute of Technology, Thane
Date:

CERTIFICATE

This is to certify that the project entitled ***"Personalized itinerary recommendation system"*** submitted by ***Nakulesh Jayakrishnan(17202012), Gauri Deshpande(16102046), Akanksha Koshti(16102011)*** for the complete fulfillment of the requirement for award of a degree ***Bachelor of Engineering*** in ***Computer Engineering***,to the University of Mumbai,is a bonafide work carried out during academic year 2019-2020.

(Prof. Sachin Takmare)
Guide

(Prof. Sachin Malave)
Head Department of Computer

Dr. Uttam Kolekar
Principal

External Examiner(s)

1.

2.

Place:A.P.Shah Institute of Technology, Thane
Date:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

_____

(Signature)

_____

(Nakulesh Jayakrishnan 17202012)
(Gauri Deshpande 16102046)
(Akanksha koshti 16102011)

Date:

Abstract

The hassle of deciding on a travel destination is often overlooked by travel websites. Travellers, a lot of times, don't have a clear idea of where they want to travel to.We aim on solving this problem by introducing a chat bot system that can recommend travel destinations based on minimal information from the traveller. Another interesting feature of the project is it's itinerary generator.The system aims on providing a human-like user experience through the use of a chat bot interface. The interface interacts with the user to retrieve information about the user's details like travel date, number of children and adults travelling and the budget of travelling.The user's budget will be the main focus of this application as we want to give the end-user the best travel experience based on their particular budget. The recommender also takes other external factors such as the season, previous traveller experiences and weather into consideration. Incorporating these factors ensure that the most optimal destination is recommended to the user.As per the recommendation, the user can opt to get several itineraries to choose from. The itinerary generator also takes several external factors into consideration when generating an itinerary. The choice of itineraries vary in the choice of places and in activities according to budget and other external factors.

Contents

List of Figures

Chapter 1

Introduction

A personalized itinerary recommendation system uses machine learning to dynamically recommend a famous tourist hotspot across India as well as curate a personalized itinerary according to the user's interest. It takes user input such as the budget of their travel, day and date of travel and number of adults and children travelling etc. Based on the interest of the user,various similar places will be recommended along with a number of personalized itineraries for that particular place, from which the user can choose an itinerary of his/her choice. The recommendation system also recommends a convenient mode of travel, keeping in mind the budget of the traveller.

1.0.1 Problem Definition

To develop a user-friendly system that recommends a travel destination based on minimum user input and also generate a personalized itinerary for the user.

1.0.2 Objectives

1. To make the process of trip planning hassle free and cost-effective.
2. To make it convenient for the users to pick a travel destination and itinerary of their choice.
3. To curate a travel plan according to the user's preferences, making the user's budget the center of attention.
4. To ensure maximum accuracy through consideration of external factors.
5. To provide the best possible user experience.

1.0.3 Scope

The system takes basic details from the user and utilizes the same along with stored external information to precisely recommend the most suitable destination in the country.The user interacts with the system through a chatbot. The system also generates a personalized itinerary for the user according to the user interest.The system is confined to tourist hotspots in India

Chapter 2

Literature Review

In the research paper 'Personalized Attraction Recommendation System' for Tourists Through Check-In Data', the authors have focused on building a personalized recommendation system that outputs the most optimal travel destination as per the user's social network data. It is most helpful in resolving the 'cold-start' problem i.e. the RS does not have any information about a new user. In such cases, the recommendation system takes the input from the social media activity of the user, since no previous data of the user is available as an input. Here, there is no need for the user to explicitly enter any data as input since the recommender system will use the user's previous data, if any, to make the recommendations suitable for the user.

The project paper on 'Crowd Prediction System for Tourists' proposes a solution for dynamically obtaining the crowd density in a particular tourist hotspot. The paper provides solutions to obtain alternate routes and analyse the overall tourist behaviour using social media data which can help optimize the itinerary generator module. This paper provides valuable insight into how we can incorporate the proposed solution into the recommendation system. The itinerary generator benefits the most from the solution proposed in this paper.

The project paper on ' Building a spatially-embedded network of tourism hotspots using geo-tagged social media data' describes the benefits of using the user's location as an input for building a network of places of tourist attraction. Constructing a network of tourism hotspots using these geotagged data would improve our understanding of tourism activities. In the hotspot network, the interconnected triplets have a tendency to be formed by the edges with greater weight values, and a high-weighted edge is often connected by two high-degree vertices. This paper provides a guide for building spatially-embedded hotspot networks based on geotagged social media data, which helps to understand the laws of travel and provides decision support for the development of tourism resources.

Chapter 3

Technology Stack

Following were the technologies used for the recommendation system:

1. Python: Python is an intercepted, high-level programming language. Python's design philosophy emphasizes code readability with its notable use of significant whitespace.

2. Jupyter Notebook: This application is mainly used for portability of the python code used in the model generation.

3. Scikit-learn: Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

4. Django framework: It is a framework for developing robust web applications.

5. jQuery: jQuery is a JavaScript library designed to simplify HTML DOM tree traversal and manipulation as well as event-handling, CSS animation and Ajax.

Chapter 4

Project Design

4.0.1 Proposed system

The main aim of our project is to make it extremely convenient for people to plan their vacations by integrating all aspects of choosing a travel destination and planning the itinerary inside one application. This means the hassle of searching multiple websites to choose a place of interest could be avoided. 'Personalized Itinerary Recommendation System' is the one-stop application that will recommend the user an array of travel destinations as well as personalized itineraries by considering various aspects such as season of travel, the budget of the user and so on. The unique thing that sets this recommendation system apart is its extremely user-friendly interface. The highlight of our application is the chatbot interface. Through the chatbot medium, the user can effectively interact with the chatbot, thereby gaining a human-like experience while communicating. With a little chatting, the user is presented with a cluster of suitable travel destinations as well as personalized itineraries right from adventure to relaxation.

4.0.2 Flow of modules

The entire project can be broadly classified into four main modules: the dataset, chatbot interface, destination recommender, and itinerary generator. The flow of modules is such that the working of the current module depends on the successful completion of the previous module. For example, the chatbot interface cannot work if the dataset module is not under working conditions or if the data is not trained accurately. The first step in the basic flow of modules includes the collection of data from relevant sources. After the data has been retrieved, the data has to be trained to generate a recommendation model. Only when this phase is successfully completed, the chatbot interface module comes into the picture. The chatbot provides a human-like interface for the user to interact with. The chatbot will take certain inputs from the user in order to recommend a place. After the chatbot is fully implemented, it can recommend a place and a personalized itinerary. While recommending a place, it will suggest a number of places of similar type. Not only that, the itinerary generator will generate a few itineraries of different types from which the user can choose according to his/her interest.

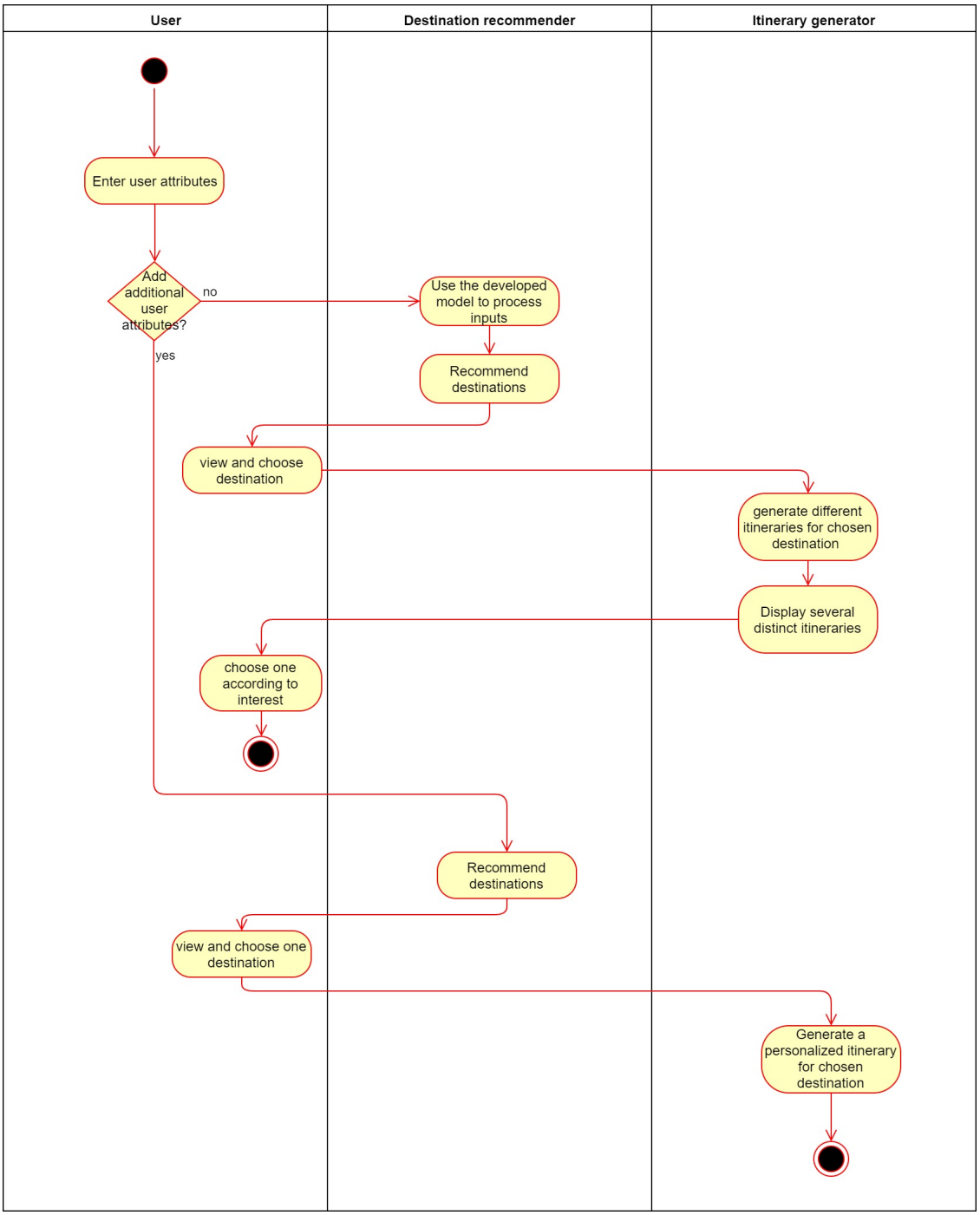

Figure 4.1: Activity Diagram

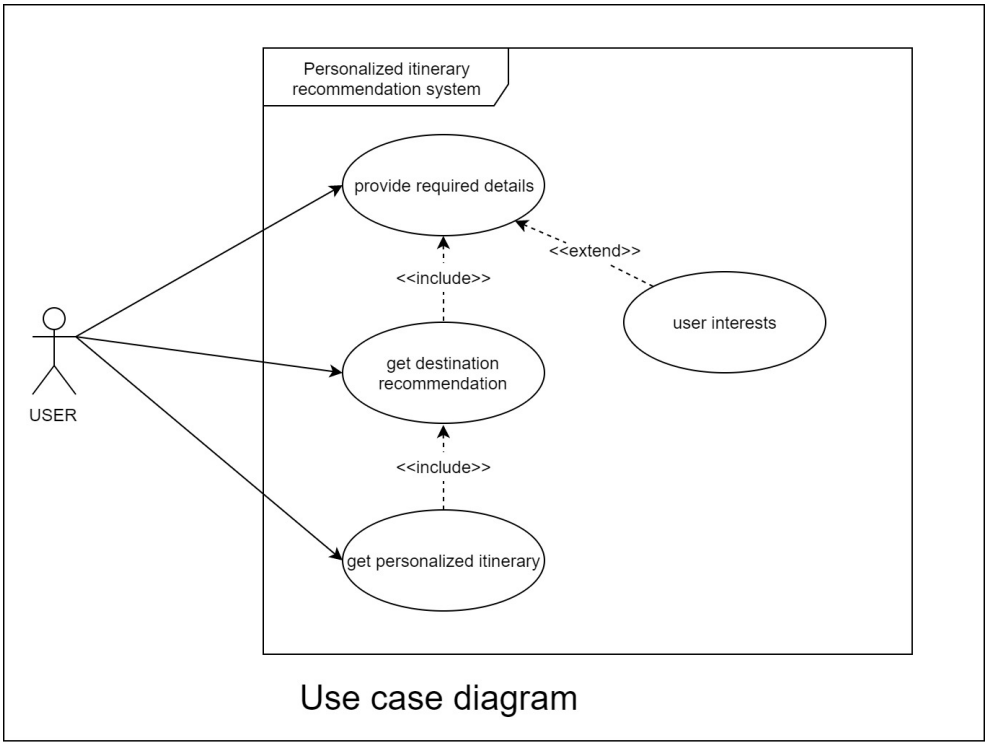

Figure 4.2: Use Case Diagram

4.0.3 Description of use case diagram

The above use case diagram shows the visual representation of the process that is required to get a travel recommendation. This use case diagram helps us to visualize the functional requirements of the system. In this diagram, the actor that is going to interact with the system is the tourist or the traveler looking for a place to travel to.

Here, the user interacts with three use cases namely; get user attributes, get travel recommendation and get a personalised itinerary. The user has to enter the desired attributes like date of travel, number of adults and children, etc. in order to proceed. A number of travel destinations would be recommended on the basis of the input attributes. Hence, the include relationship is used for dependency. For each destination, a personalized itinerary is provided according to different interests.

Therefore, personalized itinerary is dependent on the destination. 'Get user attributes' has an extension of 'user interests' wherein the user can add its interests as well and only one personalized itinerary will be provided on the basis of that. This is completely optional. In this way, the core functionalities of the system are highlighted with the help of the use case diagram.

Chapter 5

Modules of recommendation system

5.0.1 The dataset

The data is the most crucial aspect of the system since it acts as the fuel that powers the recommendation module and the itinerary generator module. The dataset has been prepared with data acquired manually from travel blogs, websites and other sources. The dataset for the recommendation system is in the form of .csv files. Attributes most relevant to recommendation of a destination are considered in the dataset.

5.0.2 Destination recommendation system

This module deals with taking the user details and processing them against a trained model to generate a cluster of recommendations to the user. This module uses the trained dataset and recommends a cluster of places to the user according to the user's interest. The input is taken from the user through the chatbot interface and the resulting recommendations are also provided through the same interface. Instead of recommending only one place, the recommendation system recommends a number of places of similar type from which the users can choose the place of their interest.

5.0.3 Itinerary generator

The itinerary generator is the last phase of the application. After the user has been recommended a cluster of places, they could choose a destination of their choice from these places, after which a personalized itinerary will be generated for the user. This itinerary will focus mainly on the interest of the user. According to the specific tags that the user will input, an itinerary catering to those specific needs of the user will be generated.

5.0.4 Chatbot UI

The chatbot interface will act as an intermediate between the user and the system. The user can interact with the chatbot in order to input the desired attributes. The interface is responsible for handling inputs to and outputs from the destination recommendation module as well as the itinerary generator module.The interface communicates with the modules, recommendation system and itinerary generator and cordinates the communication between the user and the system.

Chapter 6

System Implementation

6.0.1 Algorithms used

The two main algorithms used in the project are:TF-IDF Vectorizer and Agglomerative Clustering. The indepth analysis and use of these two algorithms are as follows:

1. TF-IDF Vectorizer:

The TF-IDF Vectorizer is predominantly used in the recommender module. TF-IDF is an abbreviation for 'Term Frequency-Inverse Document Frequency. This algorithm is used to transform text into meaningful representations of numbers. In order to use Machine Learning Algorithms, it is crucial to convert all text based data into numeric data. One of the simplest ways to represent text in the form of numbers is to count the number of times the word occurs in the entire corpus. A better way of representation would be to normalize the occurence of the words with that of the size of the documents. Numerically, Term-Frequency(TF) can be represented as:

tf(w)=doc.count(w)/total words in a document

While computing the term-frequency, each term is considered equally important. Numerically IDF can be represented as:

idf(w)=log(total number of documents/number of documents consisting word w)

2. Agglomerative Clustering:

The Agglomerative Clustering Algorithm is used in the Itinerary Generator Module. It is used to cluster the nearer places together. Agglomerative Clustering is the most common type of hierarchical clustering used to group objects in clusters according to their similarity. Hierarchical Clustering is based on the core idea of objects being more related to nearby objects than to objects farther away. This algorithm connects objects to form clusters based on their distance. There are three main parameters used in Agglomerative Clustering. They are as follows:

i. SingleLink: In SingleLink Hierarchical Clustering, we merge in each step the two clusters whose merger has the smallest distance.

ii. CompleteLink: Here, we merge in each step the two clusters whose merger has the smallest diameter.

iii. AverageLink: Average-link clustering is a compromise between the sensitivity of complete-link clustering to outliers and the tendency of single-link clustering to form long chains that do not correspond to the intuitive notion of clusters as compact, spherical objects.

6.0.2 Libraries used

The libraries used for successful implementation of the application are: Pandas and Scikit-Learn.

i. Pandas:

Pandas stands for 'Python Data Analysis Library. Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. Pandas is mainly used for machine learning in form of dataframes. Pandas allow importing data of various file formats such as csv, excel etc. Pandas allows various data manipulation operations such as groupby, join, merge, melt, concatenation as well as data cleaning features such as filling, replacing or imputing null values.

ii. Scikit-Learn:

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. The library is built upon the SciPy (Scientific Python) that must be installed before you can use scikit-learn. The library is focused on modeling data. It is not focused on loading, manipulating and summarizing data. For these features, refer to NumPy and Pandas. Scikit-learn integrates well with many other Python libraries, such as matplotlib and plotly for plotting, numpy for array vectorization, pandas dataframes, scipy, and many more.

6.0.3 Working of modules

1. NAVIGATING THROUGH THE USER INTERFACE

The User Interface is the way in which the system communicates with the user. The main aim behind keeping the User Interface precise and simple is for the user to navigate easily through the Interface. The first thing the user sees in the User Interface is the navigation tab which is placed on the topmost left hand corner. It has three simple sections called 'Home', 'About' and 'Customize my tour'. Each of the sections navigates the user to a different section in the Interface. Following are the functions of each of the section of the Interface:

A. HOME:

Home section is the current page that is under display. When the user wants to navigate to the HOME page, the 'Home' button is what they should press.

B. ABOUT:

This section of the Interface tells the user in brief about the application and its working. It is the overall explanation of the system as a whole. The 'About' section emphasizes on the importance of using the Personalized Recommendation System. It also presents the user with the three step customization process through which the user can successfully generate a destination recommendation as well as an itinerary for that specific destination.

C. CUSTOMIZE MY TOUR:

This section of the Interface is the highlight of the project. It is under this tab that the Chatbot Interface, Recommender Module and Itinerary Module lies. If and when the user wants to interact with the Chatbot Interface, the user can click on 'Customize my tour' tab and the Interface will lead him to the Chatbot Interface, where the user can interact with the system to get a personaized destination recommendation as well as a day-wise itinerary. The itinerary generator module uses the concept of agglomerative clustering. It is a type of hierarchical clustering algorithm in which the similar objects are grouped together.

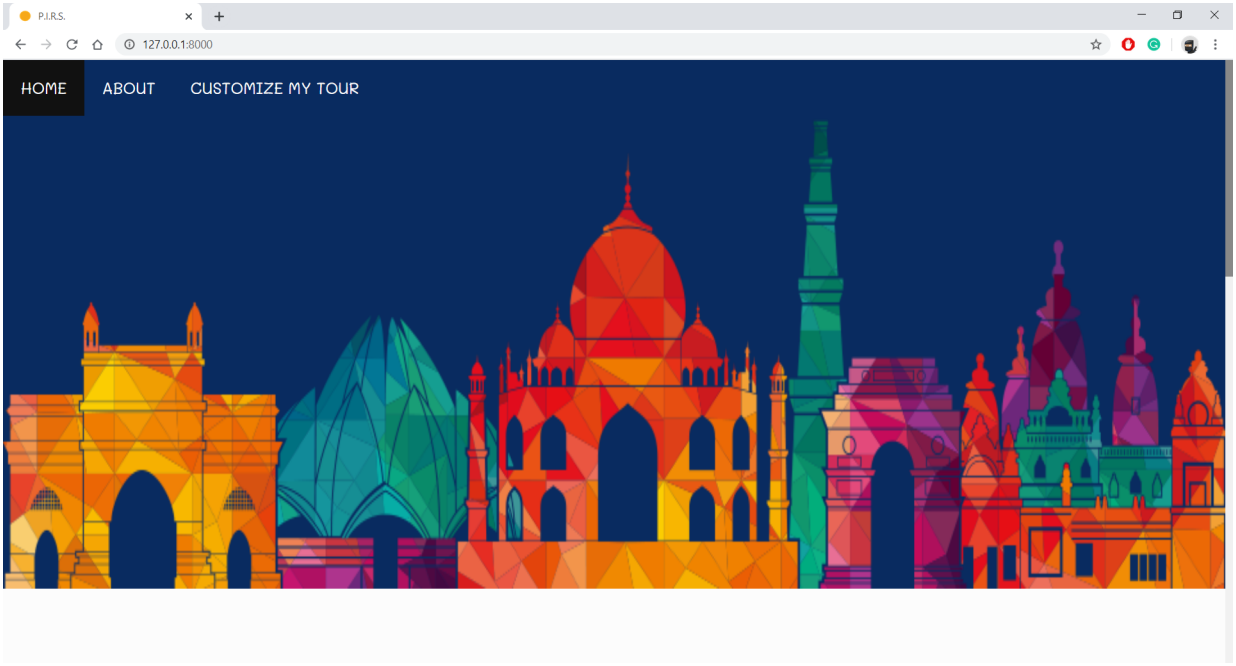

Figure 6.1: User Interface

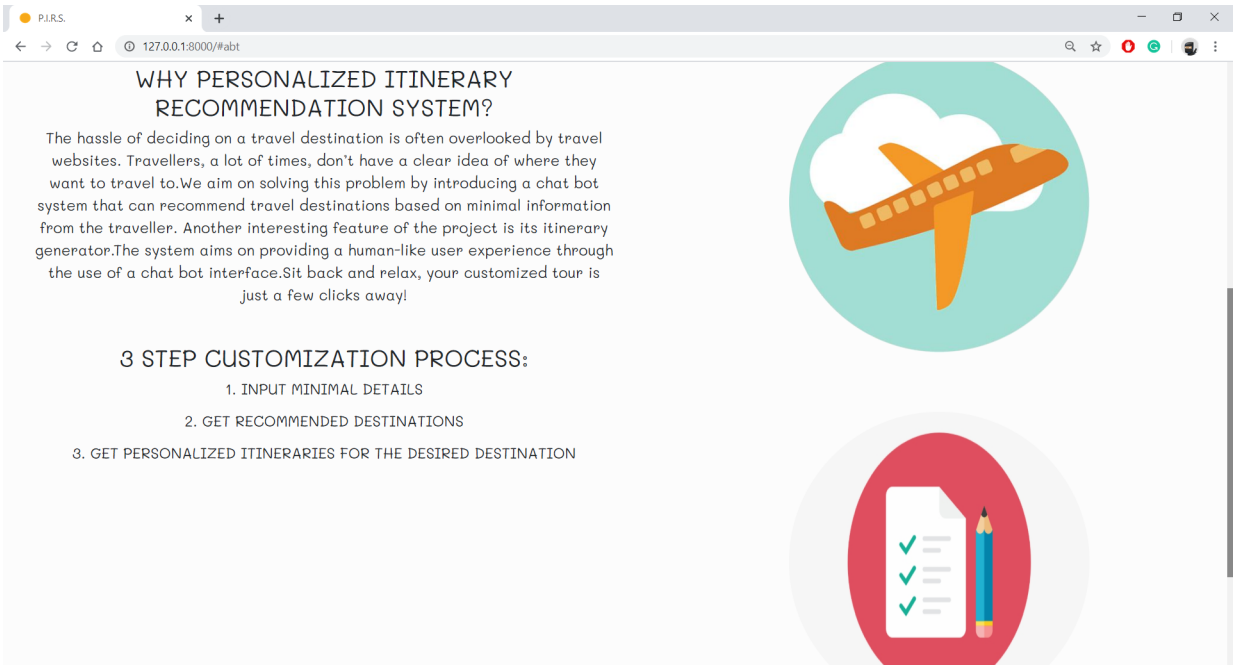

Figure 6.2: About Section

2. INTERACTING WITH THE CHATBOT INTERFACE

The Chatbot Interface is the medium of interaction between the user and the system. It is through this chatbot Interface that the user gets a recommended destination according to their interests, in addition with the personalized itinerary for that particular destination too. The chatbot interacts with the user in the following ways:

A. Asking the desired season of travel: Almost all the places in India have a specific season in which they can be explored. Choosing the right place in the right season is the key in enjoying your travel. For this purpose, the chatbot specifically asks the user the preferred season pf travel. This is one of the main attribute that is taken into consideration while choosing a destination.

B.Asking the users who they are travelling with: This is also an important factor to take into consideration while planning a trip. Based on who all are accompanying the user, the system can recommend a place suitable for that specific group. Suppose the user is looking to go on a solo trip, the system will recommend him/her destinations that are proven to be safe for solo travel.

C. Asking for additional preferences: These are the additional attributes that the user can enter, provided he/she wants to. If the user does not want to enter such preferences, the recommender module will take only the previous basic inputs and will recommend a destination based only on those previous input. However,if the user desires to enter the additional preferences, the chatbot interface will ask the user to enter the attributes such as desired terrain(mountains or beaches) and the activities that the user would be specifically interested in(adventure, relaxation or sightseeing). According to these new inputs as well as the previous ones, the recommender module will then recommend suitable destinations as well as generate itinerary for the user.

D. Getting the recommendations: Once the input attributes are entered by the user, the recommender module recommends a set of similar places to the user. The users can then choose from these recommended destinations after which the itinerary generator module will generate an itinerary that suits the needs of the user.

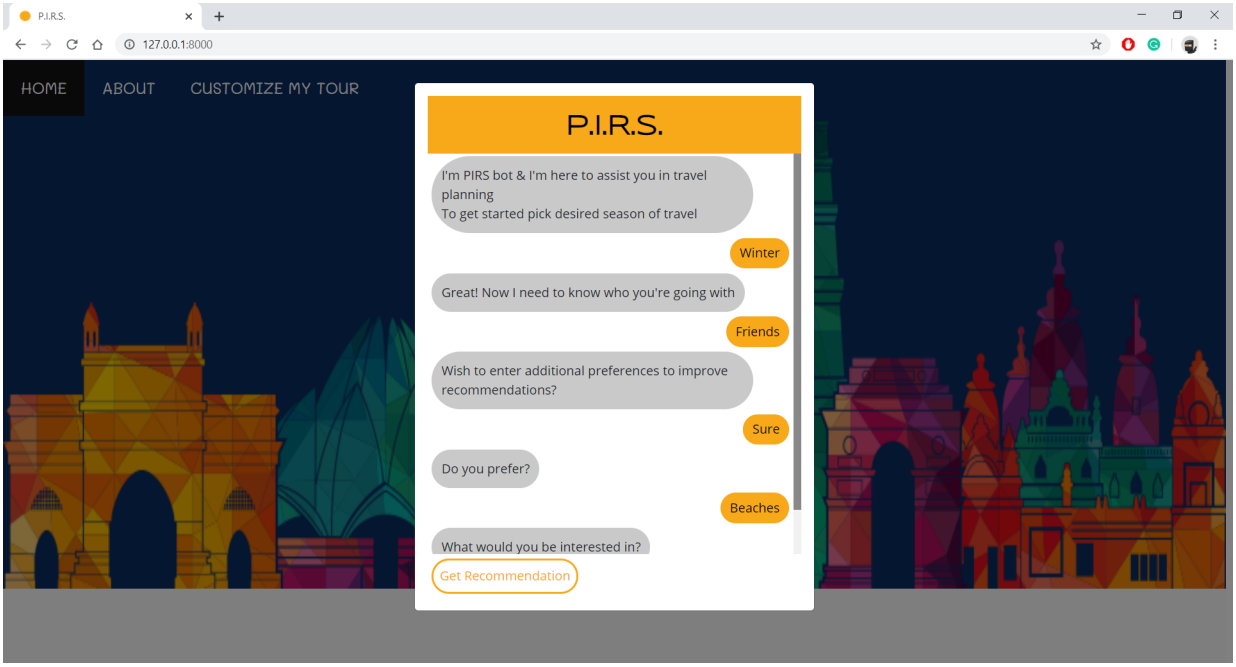

Figure 6.3: Chatbot Interface

3. GETTING RECOMMENDATIONS

After the user inputs the required attributes, the recommender module comes into play. The recommender module then recommends the user a cluster of places that are of the similar type. For example: if the user's additional preference is that of 'beaches', the recommender module recommends places like Goa, Andaman and Pondicherry. This gives the user a wide range of options to choose from. The following are the main tasks performed by the recommender module:

i. To take the user input details and to process them against a trained model to generate a cluster of places according to the user's interest.

ii. The input is then taken through the chatbot and the resulting recommendations are provided through the same interface.

iii. Instead of recommending only one place, a cluster of similar places are recommended, from which the user can pick one place of his/her desire.

4. GETTING A PERSONALIZED ITINERARY

When the user picks a specific destination from a cluster of recommended destinations, the itinerary for that particular place is generated. This itinerary is generated by the itinerary module. The itinerary consists of the places that are famous tourist hotspots of that destination. The itinerary is presented in a day-wise format. This makes it easier for the user to plan their holiday. The places which are nearer to each other or in the vicinity of each other are scheduled on the same day. This ensures that the user is not spending maximum time in travelling from one place to another. The itinerary generation is the last phase of the entire application. The motive behind generating an itinerary is to help the user smoothly plan their course of vacation without much hassle. Having an itinerary ready can help the user to maximise his/her travel experience.

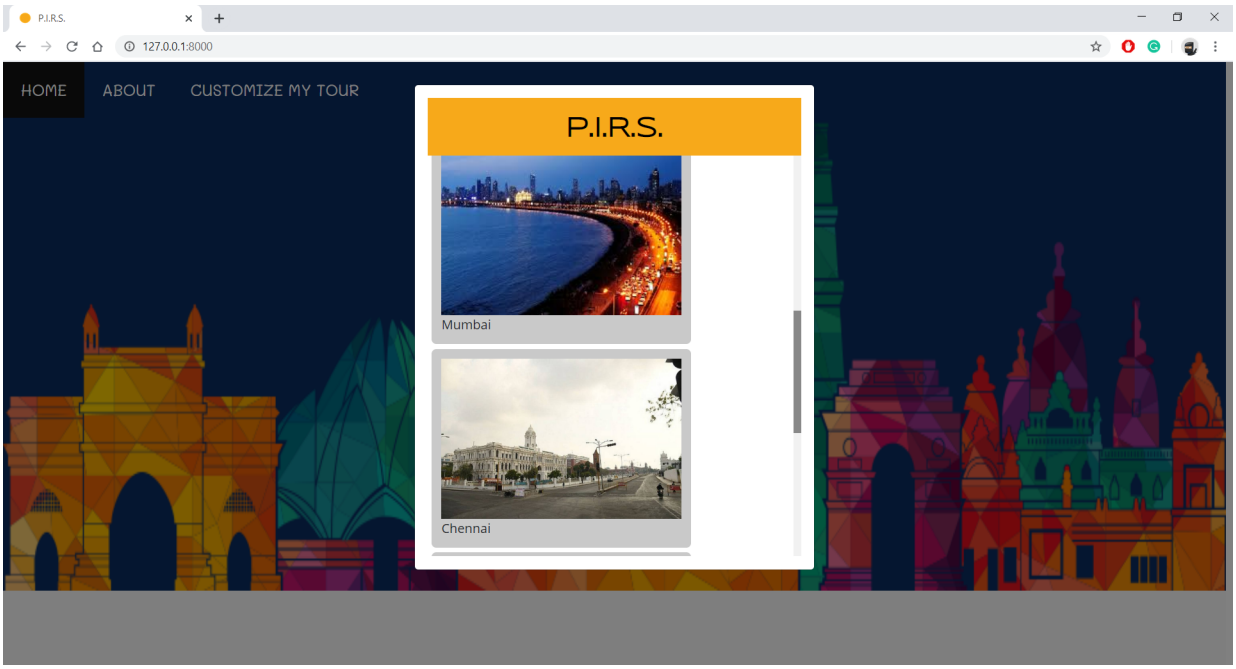

Figure 6.4: Recommendations

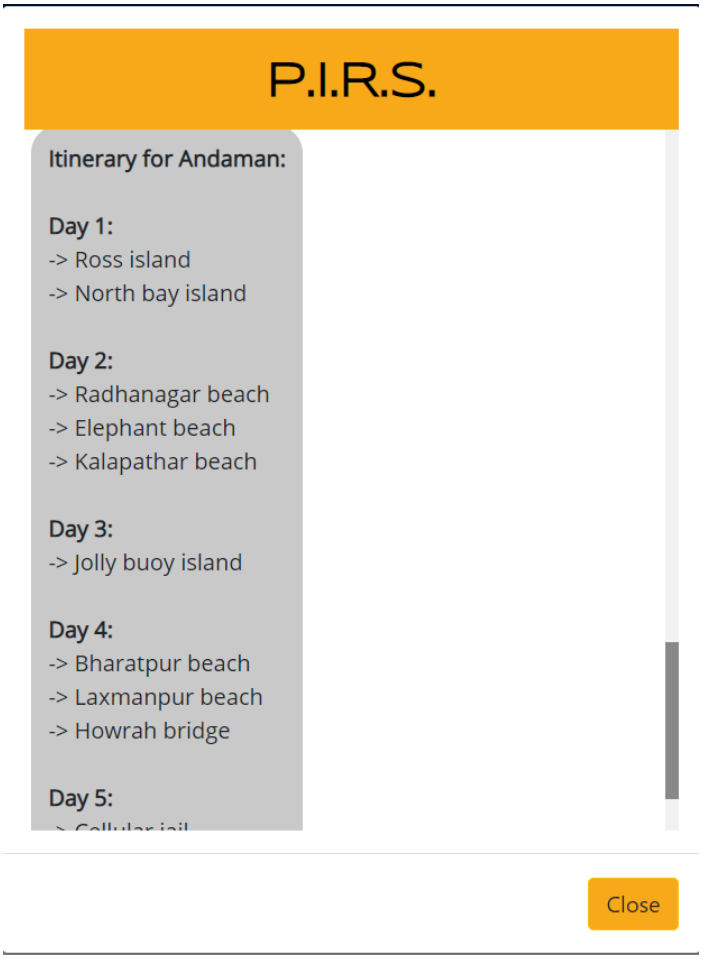

Figure 6.5: Generated Itinerary

Chapter 7

Benefits for environment and society

Instead of physically going to the travel agencies for the recommendation, the users can simply interact with the personalized itinerary recommendation system through a chat bot. Moreover, the travel agencies provide information to the users through brochures or pamphlets, resulting in a huge amount of paper wastage.

Tourism is the fastest growing economic sector in terms of foreign exchange earnings and job creation. It has become one of the most important sources of employment. It also provides financial support for the conservation of the ecosystem and natural resources management making the destination more reliable and desirable to visitors.

Easy interaction that the system provides to the user through a chat bot, it attracts more and more users to use this system thereby attracting them to different places. Certain users may even leave a full review of the places that they have visited which in turn benefits the other users. This helps to increase tourism and brings people into closer contact with nature and the environment.

Chapter 8

Result

The two main concepts used to get the required results were: content-based filtering and cosine similarity. The places of similar kind are clustered together by giving each place various tags according to the relevance of that place. These tags are then mapped according to the user's input and the recommended places are given as an output based on cosine similarity.

Cosine similarity is a metric used to measure how similar the documents are irrespective of their size. Mathematically, it measures the cosine of the angle between two vectors projected in a multi-dimensional space. The cosine similarity is advantageous because even if the two similar documents are far apart by the Euclidean distance (due to the size of the document), chances are they may still be oriented closer together. The smaller the angle, higher the cosine similarity.

Cosine similarity basically measures the similarity between two sentences, which in this case are the tags. When a particular tag is matching or is similar to the attribute entered by the user, it is considered for the output result. The result thus, is a cluster of places of similar kind as per the user's interest.

Chapter 9

Conclusion and future scope

It can be concluded that the recommendation system is an efficient way to provide the end user with quick, precise results. The system makes it easier for a naive user to navigate through the given options of the recommended similar places and choose a place of his/her liking.

The future scope of this project includes creating a personalized itinerary for a particular place of interest that the user decides to choose. Here, not just one, but various itineraries will be generated for the user to choose from, if the user chooses to not explicitly enter the optional attributes. However, if the user decides to enter the additional attributes, a fixed personalized itinerary will be generated according to the user's specific interest.

This will thus save the user from the hassle of having to choose a perfect itinerary for the desired destination. The whole purpose of itinerary generation is to make it extremely convenient for the user to plan a holiday without having to look explicitly on a number of various websites. This creates a one-stop platform for the user, making the whole concept user-centric.

Bibliography

[1] K. Kesorn, W. Juraphanthong and A. Salaiwarakul, "Personalized Attraction Recommendation System for Tourists Through Check-In Data," in IEEE Access, vol. 5, pp. 26703-26721, 2017.

[2] X. Wu, Z. Huang, X. Peng, Y. Chen and Y. Liu, "Building a Spatially-Embedded Network of Tourism Hotspots From Geotagged Social Media Data," in IEEE Access, vol. 6, pp. 21945-21955, 2018. pp. 26703-26721, 2017.

[3] Kartik Rajendra Kokane, Pratik Pradeep Jogdand and Rohit Dhuri, "Crowd Prediction System for Tourists"

Appendices

APPENDIX-A: Download and Installation of Python

1. Open a browser window and navigate to the Download page for Windows at python.org.
2. Underneath the heading at the top that says Python Releases for Windows, click on the link for the Latest Python 3 Release - Python (As of this writing, the latest is Python 3.6.5.)
3. Scroll to the bottom and select either Windows x86-64 executable installer for 64-bit or Windows x86 executable installer for 32-bit.

APPENDIX-B: Download and Installation of Jupyter Notebook:

1. Download Anaconda. Download Anaconda's latest Python 3 version (currently Python 3.7).
2. Install the version of Anaconda which you downloaded, following the instructions on the download page.
3. Jupyter Notebook has been installed. To run the notebook type: jupyter notebook.

Acknowledgement

We have great pleasure in presenting the report on **Personalized Itinerary Recommendation System.** We take this opportunity to express our sincere thanks towards our guide **Prof. Sachin Takmare** Department of COMPUTER, APSIT Thane for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards his constant encouragement, support and guidance through the development of project.

We thank **Prof. Sachin Malave** Head of Department, COMPUTER, APSIT for his encouragement during progress meeting and providing guidelines to write this report.

We thank **Prof. Amol Kalugade** BE project co-ordinator, Department of COMPUTER, APSIT for being encouraging throughout the course and for guidance.

We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement.

Student Name:Nakulesh Jayakrishnan
Student ID:17202012

Student Name:Gauri Deshpande
Student ID:16102046

Student Name:Akanksha Koshti
Student ID:16102011