Accomodation Recommendation and Booking for Students

Sakshi Rane[1], Shraddha Deskhmukh[2], Nidhee Agarwal[3], Aditya Verma[4], Dr. Amol Dhakne[5]
Associate Professor[5]
Computer Department, D Y Patil Institute of Engineering
Management and Research Pune, India

Abstract:- Students in Various colleges travel from far away places for a better education. Here they often face problem while looking for an affordable and safe accommodation. Additionally, users often have to check out the places as there is very little information in amenities , location and user feedback. Not only this but also often, people are too tired to prepare home-cooked meals due to the fast-paced and busy environment where they live. Furthermore, even if you eat only homemade food every day, you are still likely to want to go out to eat for social or recreational purposes from time to time. Despite this, it's common knowledge that food is an important aspect of anyone's lifestyle, no matter where they live. Assume, for example, someone has just moved to a new home. Their preferences and tastes are already established. If a student lives nearby her favourite outlet, this will save her a lot of trouble and help her save money. Students nowadays also have to take care of their health and fitness. Hence, recommendation for affordable gyms has also become a requirement. Based on incoming students' preferences for facilities, budget, and proximity to the location, this project uses K-Means Clustering to find the most suitable accommodation, restaurant as well as gyms for them in Akurdi (Pune).

Keywords:- Geolocation data analysis, k- means clustering, python, recommendation.

I. INTRODUCTION

The majority of Indian migrants migrate for study purposes, with 2.9 million moving within the state from their previous residence for less than 5 years. There are around 50,000 registered hostels and accommodation services available in India.

Geolocational data analysis is the process for analysing data based geographic location. It is possible to make recommendations for accommodations based on the location of the user and the location of available accommodations through the use of geolocational data analysis. Using this type of analysis it can also be used to identify various trends and patterns in the behaviour and preferences of the user. Overall, geolocation data analysis can significantly improve the accuracy rate of our proposed system.

In our proposed system we are using technologies that provide maximum benefit and usability regarding geolocational data analysis. For front-end we are making use of REACT, it is a front-end JavaScript library. We are using it to make our website's interface reactive and more userfriendly. React is capable of making API calls (sending the request to the backend), which deal with the data. For backend the best could be NodeJs, and database management system we have used Mongo DB, For API we have used insomnia.

The methodology used here is k-means clustering. It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabelled dataset on its own without the need for any training. It is a clustering algorithm that helps in forming clusters and classification of data in various groups.

II. EASE OF USE

This is the software configuration in which the project was shaped. The programming language used, tools used, etc are described here.
- Python
- REACT JS
- Node JS
- Database management system.
- For data - numpy and pandas package. 15 • For plotting - matplotlib package & seaborn packages
- For geospatial - geopy, folium.
- •For machine learning – sklearn, scipy

A. back-end
This is for the working of backend development which is configuration, administration and management of databases and servers
- NodeJs- Create a server side web application
- MongoDB- Used as a backend data store
- Insomnia- Open source application For the api connection, designing, debugging and testing
- REST

III. METHODOLOGIES

- **K-Means Clustering**

A. Introduction
K-Means Clustering is an unsupervised learning technique that groups unlabelled data into distinct clusters. Here K defines the number of groups that should be created in this process, if K=2 there will be 2 groups, if K=3 there will be 3 groups and so on. It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabelled dataset on its own without the need for any training. The main objective of the K-Means algorithm is to minimize the sum of distances between the points and their respective cluster centroid. This algorithm is based on centroids. Each cluster is assigned a

centroid and tries to minimize the distance between each point to cluster. Here our data points are the locations of particular place i.e. the accommodation, hostel, hotel, restaurant, mess, gyms, library etc. and the clusters so formed are according to the filters applied by the user according to their preferences and choices.

B. Steps in k-means clustering

➢ *Choose the number of clusters k*
The first step in k-means is to pick the number of clusters, k. 2. Select k random points from the data as centroids

Next, we randomly select the centroid for each cluster. Here, the red and green circles represent the centroid for these clusters.
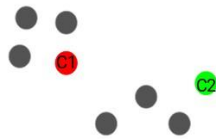
➢ *Assign each point to the nearest cluster centroid.*
Every point is assigned to the nearest cluster centroid. The points closest to the red point are assigned to the red cluster, while the ones closest to the green point are assigned to the green cluster.
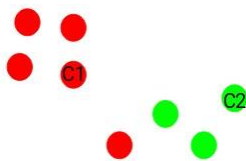
➢ *Calculate the centroids of freshly generated clusters again.*
In the next step, we will compute the centroids of the newly created clusters after all of the points have been allocated to either cluster. Centroids have been revised as indicated by the red and green crosses.

➢ *Repeat steps 3 and 4*
We then repeat steps 3 and 4:
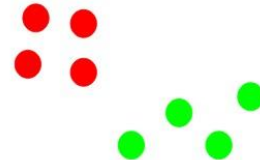
➢ *Stopping criteria for k-means clustering*
There are essentially three stopping criteria that can be adopted to stop the K-means algorithm:
- Centroids of newly formed clusters do not change
- Points remain in the same cluster
- Maximum number of iterations is reached

IV. IMPLEMENTATION

K-means clustering or any other algorithm in machine learning can be implemented by either using Jupyter Notebook or Google Collaboratory or Colab which provides us with necessary interface and runtime. These are opensource web-based interactive computing platforms; Jupyter Notebook is an incredibly powerful tool for interactively developing and presenting data science projects. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education.

A. BASIC PROCEDURE
Implementing k-means clustering algorithm in python needs a few libraries to be installed first.

➢ *So the first step is to install the necessary libraries needed*
To be able to implement k-means we have libraries like sklearn, cluster, numpy, pandas, etc.

```
Run KMeans Clustering on the data

[ ]  ## for data
     import numpy as np
     import pandas as pd
     ## for plotting
     import matplotlib.pyplot as plt
     import seaborn as sns
     ## for geospatial
     import folium
     import geopy
     ## for machine learning
     from sklearn import preprocessing, cluster
     import scipy
     ## for deep learning
     import minisom
```

Fig. 1: Installing libraries

➢ *Select the dataset to be used*
We find the appropriate dataset that is to be used for evaluation and finding answers. We then upload this dataset and begin with data exploration.

➢ *Pre-processing of data*
Once we have the desired the data, it still contains some unwanted information, we remove this by pre-processing and cleaning the data.

➢ *Data visualization*
We use various plotting techniques to visualize our data before applying the algorithm to get an overview of the dataset.
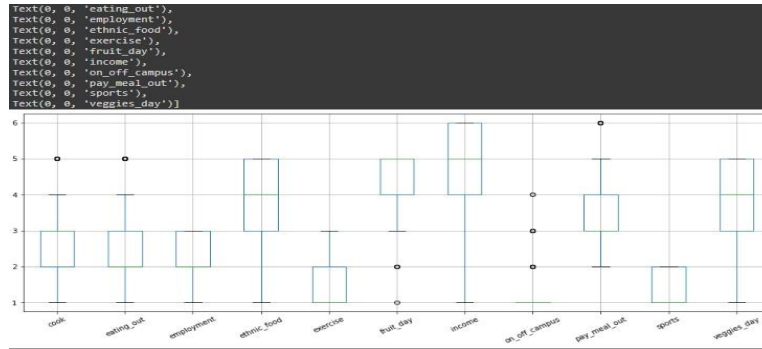

Fig. 2: Data Visualization

➢ *Geolocational data*
We then using the API collect and feed the geolocational inputs to the model and expect an output.

Fig. 3: Geolocational data

B. *Application Programming Interface(API)*

➢ *INTRODUCTION*
In many places, geolocation services are becoming essential for a variety of applications and disciplines. When the consumer is Many websites that are frequently visited use geolocation data for targeted localized advertising, localized content (such as local news and weather), and adherence with local laws occasionally without the user being aware of it. APIs play a key part in enabling these functions. In the context of API's, the word Application refers to any software with a distinct function. Interface can be thought of as a contract of service between two applications. This defines how the two communicate with each other using requests and responses. API automation is used as API's are the internal connections which are present on both web and mobile application, which is hidden from the user, but critical for the application to perform functions. It is analogous to human body containing blood vessels and nervous system. Hence when performing automation in this layer it requires specific targeted information, inputs and outputs. API- in this project

mainly apo's are inherited from the Foursquare, which is mostly used for building the review app structure or ride sharing/navigation to the user. Third-party apps can connect to Foursquare and augment the official app's functionality. Developers can gain access to and collect useful data for analysis. The input given to this model is the current location and the radius of the search for which you need the recommended places. And the output gives all the popular places around a given location and python to visualize this stuff.

➢ *STEPS IN API IMPLEMENTATION USING FOURSQUARE API*
• **Data Requirement:** Visit the foursquare developer website and a developer account needs to be created Client id and Client secret credentials will be provided .Now to explore a location we use this URL where latitude and longitude are current coordinates which we can get using geopy library. Limit is the number of results which will be fetched.

```
https://api.foursquare.com/v2/venues/explore?client_id={CLIENT
ID}&client_secret={CLIENT SECRET}&ll={Latitude},{Longitude}&v=
{VERSION}&radius={RADIUS}&limit={LIMIT}
```

Fig. 4: URL of coordinates

- Model Development: 1)Importing all essential python libraries 2)Defining the foursquare API credentials. 3)Read the current location from the input and converting it to Coordinates.

Get Geolocational Data

```
from pandas.io.json import json_normalize
import folium
from geopy.geocoders import Nominatim
import requests
CLIENT_ID = "KTCJJ2YZ2143QHEZ2JAQS4FJIO5DLSDO0YN4YBXPMI5NKTEF" # your Foursquare ID
CLIENT_SECRET = "KNG2LO22BPLHN1E3OAHWLYQ5PQBN14XYZMEMAS0CPJEJKOTR" # your Foursquare Secret
VERSION = '20200316'
LIMIT = 10000
```

```
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    17.448372, 78.526957,
    30000,
    LIMIT)
```

```
results = requests.get(url).json()
```

Fig. 5: Model Development

- Fetching data from the foursquare API, returns output as JSON data.

Adding two more Columns Restaurant and Others

1. Restaurant: Number of Restaurant in the radius of 20 km
2. others:Number of Gyms, Parks,etc in the radius of 20 km

```
resta=[]
oth=[]
for lat,long in zip(nearby_venues['venue.location.lat'],nearby_venues['venue.location.lng']):
    url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
        CLIENT_ID,
        CLIENT_SECRET,
        VERSION,
        lat,long,
        1000,
        100)
    res = requests.get(url).json()
    venue = res['response']['groups'][0]['items']
    nearby_venue = json_normalize(venue)
    df=nearby_venue['venue.categories']

    g=[]
    for i in range(0,df.size):
        g.append(df[i][0]['icon']['prefix'].find('food'))
    co=0
    for i in g:
        if i>1:
            co+=1
    resta.append(co)
    oth.append(len(g)-co)

nearby_venues['restaurant']=resta
nearby_venues['others']=oth
nearby_venues
```

Fig. 6:

- Extracting the relevant data from Json which fetches the raw data.
- Cleaning the data hence converting it into data frames, so its ready and visualization takes place

V. PROPSED SYSTEM

The proposed system fosters hostels, apartments, and houses, and it additionally gives data on those houses, flats, and hostels (for example, images and contact information).

It suggests budget-friendly lodging. It has a lot of houses on our budget. It also has the potential in the future

to propose restaurants, gyms, and other services depending on the user's budget and tastes. It delivers accurate recommendations with no gaps. In this project, we are employing the K-means technique, although it has a downside when two circular clusters with the same mean have different radii. K-Means defines a cluster's median using median values and does not distinguish between the two clusters. When the sets are noncircular, it likewise fails.
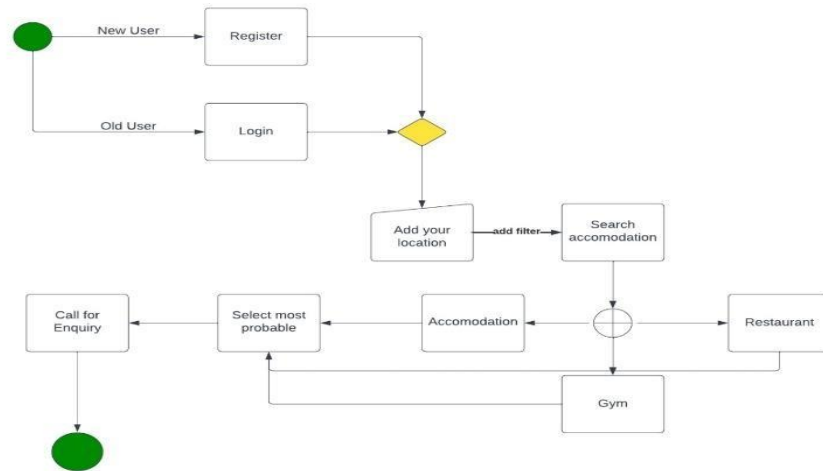

Fig. 7: UML data

VI. CONCLUSION

The similar existing models provide a few facilities that are not sufficient, some newly added features in our model helps students in not only finding the best accommodation but also other places like mess, restaurants, gyms, library etc. The use of k-means clustering helps in finding the appropriate location according to filters and choices, this technique can cluster the migrant data and geolocational data. Our website represents the analysed model in form of a plotted-map. This website will prove to be best suited for students to ease their shifting process to a new city with the recommendations filtered according to their interests. Future research will focus on taking multiple locations but not a fixed location and applying clustering techniques on it.

ACKNOWLEDGMENT

REFERENCES

[1.] S. -H. Wu, J. Hsiao, Y. -S. Wu and J. -T. Jeng, "AI Food Recommendation Systems," 2022 IET International Conference on Engineering Technologies and Applications (IET-ICETA), Changhua, Taiwan, 2022, pp. 1-2, doi: 10.1109/IET-ICETA56553.2022.9971598.

[2.] P. T. Chiou and G. S. Young, "Implementing Recommendations of Accessibility Technology Guidelines — The Quantitative Effects and Benefits it Offers to Non-disabled Students," 2017 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2017, pp. 1137-1142, doi: 10.1109/CSCI.2017.198.

[3.] M. N. Y. Utomo, T. B. Adji and I. Ardiyanto, "Geolocation prediction in social media data using text analysis: A review," 2018 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2018, pp. 84-89, doi: 10.1109/ICOIACT.2018.8350674.

[4.] S. R. Kishan, K. S. Kakunuri, A. Raj Parasa, V. Patlolla and P. Gamini, "Exploratory Analysis on Geo-Locational Data," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 1-4, doi: 10.1109/ICACRS55517.2022.10029164.

[5.] M. N. Y. Utomo, T. B. Adji and I. Ardiyanto, "Geolocation prediction in social media data using text analysis: A review," 2018 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2018, pp. 84-89, doi: 10.1109/ICOIACT.2018.8350674.

[6.] Singh R, Manchanda N, Mishra R. Internal student migration in India:

[7.] Impact of the COVID-19 crisis. Asian Pac Migr J. 2022 Dec;31(4):454–77. doi: 10.1177/01171968231154590. Epub 2023 Feb

[8.] PMCID: PMC9944466.