JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ACCOMMODATION SUGGESTION BASED ON GEOLOCATIONAL DATA USING MACHINE LEARNING

Mrs. Y.Sowjanya [1], B.Vishal[2], E.Sai Manideep[3], A.Naveen Kumar[4]

[1]Assistant Professor, Department of Information Technology Anurag University, Hyderabad, India

[2,3,4]Students, Department of Information Technology Anurag University, Hyderabad, India

ABTRACT : - An abstract for accommodation based on geolocation data could describe the use of location-based technologies and data to help travelers find suitable accommodation options. By analyzing geolocation data, such as the user's current location, preferred travel destinations, and past travel behavior, a travel platform could provide personalized recommendations for hotels, vacation rentals, and other types of accommodations that meet the user's needs and preferences. This project uses K-Means Clustering to find the best accommodation for students by categorizing accommodation based on their preferences for facility, budget, and proximity to the location. To retrieve, clean, analyze, and run K Means Clustering on geo-located data in order to recommend accommodations to incoming students in a city.

Keywords : K-Means,Location,data,dataset,Map.

I. INTRODUCTION

This project involves the use of K-Means Clustering to find the best accommodation for students by classifying accommodation for incoming students on the basis of their preferences on amenities, budget and proximity to the location.[1]

The project's implementation will take you through a day in the life of a data science engineer, from gathering data from real-world datasets, to visualising the data, to executing machine learning algorithms, to presenting the findings.[2] Managers of restaurant chains and hotels can also benefit from this information, in addition to food delivery apps. For instance, if a restaurant management is already familiar with the demographics of his existing

client he would ideally like to operate in a site where this group is concentrated in the greatest numbers, ensuring quick access to the establishment and a greater number of clients serviced.[3] In order to ensure that every visitor has something they will enjoy, it is ideal when considering potential hotel locations to choose a place that offers a wide range of cuisines.

This project is a good start for beginners and a refresher for professionals who have dabbled in python / ML before. [4] The methodology can be applied to any location of one's choosing, so feel free to innovate!

II. LITERATURE REVIEW

Mining Geolocation Data for Disease Surveillance: A Review" by C. Yang, et al. (2018)

This paper explores the use of geolocation data for disease surveillance. The authors review various approaches for mining geolocation data and highlight the challenges involved in this process. They also discuss the ethical issues surrounding the use of geolocation data in disease surveillance.[4]

"Exploring Geolocation Data for Social Media Analytics: A Review" by M. T. Islam, et al. (2019)

This paper reviews the use of geolocation data for smart mobility applications, such as traffic management, public transportation, and car-sharing services. The authors discuss the challenges involved in processing and analyzing large amounts of geolocation data and provide recommendations for future research in this area.[5]

"Exploring Geolocation Data for Smart Mobility Applications: A Review" by S. Islam, et al. (2020)

This paper reviews the use of geolocation data for social media analytics. The authors highlight the potential benefits of using geolocation data in social media analytics, such as identifying trends and patterns in user behavior. They also discuss the challenges involved in processing and analyzing geolocation data from social media.[6]

"Exploring Geolocation Data for Tourism and Hospitality: A Review" by C. Li, et al. (2019)

This paper explores the use of geolocation data for tourism and hospitality applications. The authors discuss the potential benefits of using geolocation data in this context, such as personalized recommendations for tourists. They also discuss the challenges involved in collecting and analyzing geolocation data in the tourism and hospitality industry.[7]

"Exploring Geolocation Data for Crime Analysis: A Review" by M. Hussain, et al. (2021)

This paper reviews the use of geolocation data for crime analysis. The authors discuss the potential benefits of using geolocation data in crime analysis, such as predicting crime hotspots and identifying crime patterns. They also discuss the challenges involved in processing and analyzing large amounts of geolocation data for crime analysis.[8]

III. DATA COLLECTION

To answer specified research questions, test hypotheses, and assess results, data collection is the act of acquiring and measuring information on variables of interest in a systematic and defined manner. All academic disciplines, including the humanities, social sciences, business, and natural and applied sciences, share the data collection component of research. Although techniques differ depending on the profession, the importance of ensuring accurate and truthful collection does not change. The practice of correcting or deleting inaccurate, damaged, improperly formatted, duplicate, or incomplete data from a dataset is known as data cleaning. There are numerous ways for data to be duplicated or incorrectly categorized when merging multiple data sources. Even if results and algorithms appear to be correct, they are unreliable if the data is inaccurate. There isn't just one way to prescribe something. There is no one absolute way to prescribe the exact steps in the data cleaning process because the processes will vary from dataset to dataset. But it is crucial to establish a template for your data cleaning process so you know you are doing it the right way every time.

Gender	breakfast	cook	cuisine	diet_curre	drink	eating_ch	eating_ou	employme	ethnic_foc	exercise	father_ed	fav_cuisin	fav_food	fries	fruit_day	grade_lev	ideal_diet	income	i
2	1	2	6	1	1	1	3	3	1	1	5	3	1	2	5	2	8	5	
1	1	3	1	2	2	1	2	2	4	1	2	1	1	1	4	4	3	4	
1	1	1	3	3	1	1	2	3	5	2	2	1	3	1	5	3	6	6	
1	1	2	2	2	2	1	2	3	5	3	2	3	1	2	4	4	2	6	
1	1	1	2	2	2	3	2	2	4	1	4	1	3	1	4	4	2	6	
1	1	3	6	2	2	1	1	3	4	2	1	6	3	1	2	2	2	1	
2	1	2	1	3	1	2	2	3	5	1	4	4	1	1	4	4	2	4	
1	1	3	1	1	2	2	2	2	2	2	3	5	1	1	5	2	2	5	
1	1	3	1	1	1	2	5	2	5	5	5	1	3	1	4	1	6	5	
1	1	3	1	1	1	1	3	3	5	1	5	1	1	1	5	1	2	4	
1	1	1	1	1	2	3	2	1	5	1	2	4	1	1	5	3	7	3	
1	1	3	1	1	1	4	1	2	5	1	3	1	1	1	5	2	2	5	
2	1	5	1	1	2	2	1	3	4	3	3	4	3	1	4	1	1	5	
1	1	2	1	2	2	1	4	2	5	2	2	1	1	1	5	3	2	5	
2	1	3	1	1	2	3	2	3	4	2	4	4	3	1	5	3	1	5	
2	2	4	1	2	2	3	4	3	1	1	5	1	2	1	3	1	2	4	
1	1	3	6	1	1	2	1	2	2	2	1	7	1	1	5	1	2	1	
2	1	3	1	3	2	1	2	3	4	1	2	5	3	1	3	1	1	6	
1	1	3	1	1	1	4	3	2	1	3	5	5	3	1	5	2	2	5	
2	1	4	1	2	2	3	2	3	3	2	5	1	1	1	2	1	3	5	
2	1	5	1	2	2	1	4	2	4	1	5	2	3	1	2	2	1	6	
2	1	3	3	2	2	1	1	2	3	1	4	8	2	1	3	4	7	6	
2	1	5	2	1	1	1	1	2	3	1	4	1	1	2	4	1	2	6	

Figure – I(Dataset)

IV.SEARCHING MODULES

The search module lets users search for specific content on your site. You can search both for users and for particular words. When you are on the "Content" tab of Search, you will be able to search for words appearing in the default rendering of node content on your site, which would include the default rendering of any CCK fields, Location fields, Taxonomy, etc., as well as comments. When you are on the "Users" tab of Search, you will be able to search the user names of registered users on your site, and if you have sufficient permissions, also their email addresses.

Step 1: Specify the number of clusters. In the first step we need to specify the number of clusters.

Traditionally researchers will conduct analysis multiple times, exploring different numbers of clusters.

Step 2: Allocate objects to clusters. The simplest approach is to randomly assign objects to the cluster, but there are many

other approaches (for example, using hierarchical clustering).

Step 3: Allocate each observation to the closest cluster center. After assigning the object to the cluster, we analyzed

that some points were close to the location. Matching the observations to the closest cluster yields the following graph. When we reallocate the observations to the closest clusters

V.PROPOSED METHODOLOGY

1. Fetch datasets from the relevant locations. (Data Collection)
2. Clean the datasets to prepare them for analysis. (Data Cleaning via Pandas)
3. Visualize the data using boxplots. (Using Pandas/Tkinter)
4. Fetch geolocational data Hear API. (REST APIs)
5. Use K-Means clustering to cluster the locations. (Using Scikit Learn)
6. Present findings on a map. (Using HEAR API/Seaborn)

VI. Work Flow of the Project

Suitable parameters are chosen to perform data cleaning. The cleaned data is visualized using Box-plots. Cleaned Data Set

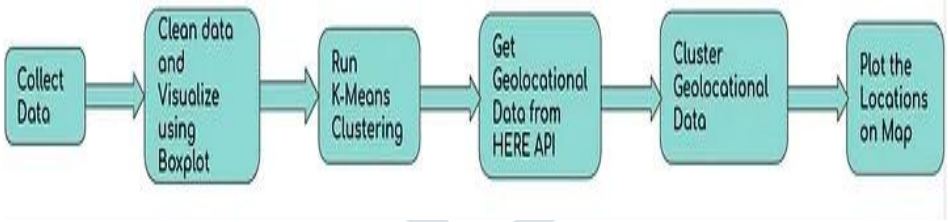
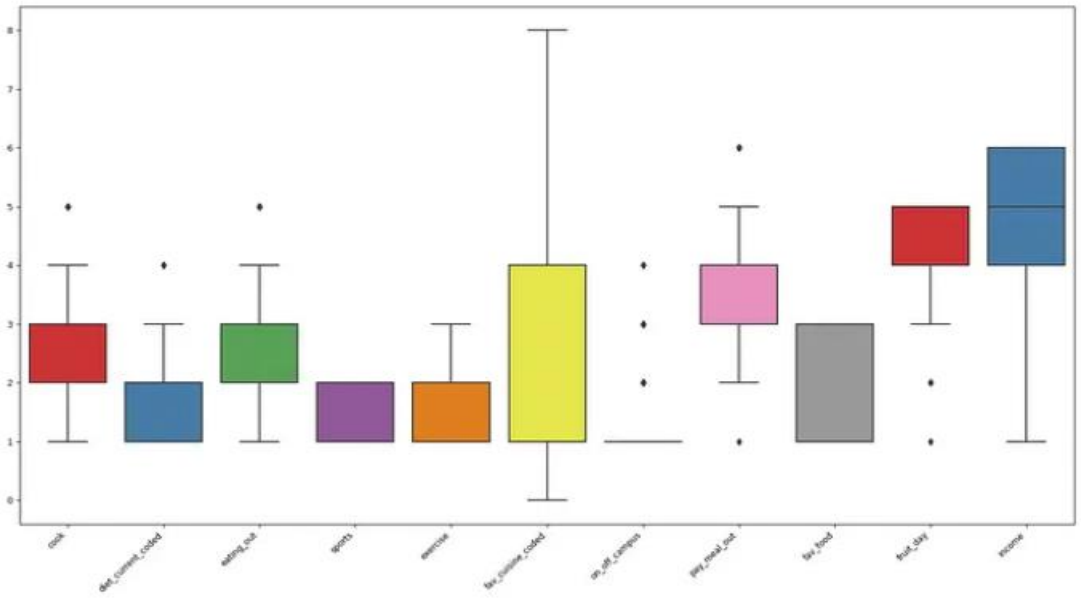


Figure – II

Box-Plot for Cleaned Data

- K-Means clustering is applied on cleaned data for arbitrary values of K and best value of K is found.[1]
- Box-Plot for optimal K K=3
- for K=2 : no clear demarcation is seen between the respective attributes
- for K=3: as we can see in the above box-plot, a clear demarcation on the basis of income can be seen.



- **High income students :** In general they pay more on meal out, eat less fruits, mostly stay on campus, exercise more, cook less.
- **Low income students :** In general they pay less on meal out, eat more fruits, more likely to stay away from campus, exercise less, cook more often.
- However the boundaries of these classifications are not very prominent.
- Now we use real time data from HEAR API (HERE Geo coding API) for any to find accommodation for students around the area.
- Useful parameters are selected an data cleaning is performed on this data.
- Cleaned API data
- This data is now clustered using optimal K found in step 3.
- Clustered API using K=3
- Using Hear Api , the above clustered data is plotted on map.

Cluster 0(Orange): First Preference

Cluster 1(Yellow): Second Preference

Cluster 2(Green): third preference

VII.RESULTS

USERSCREEN

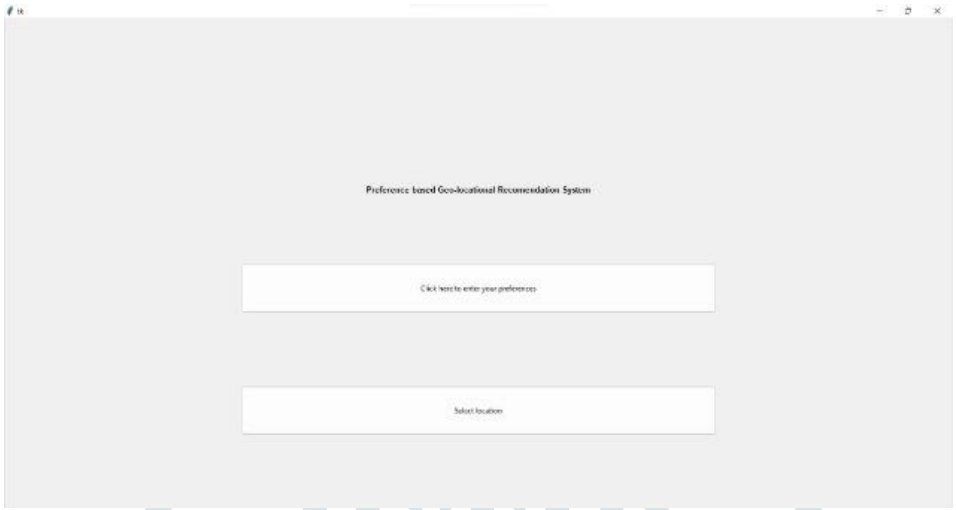


Figure-III

PREFERENCES PAGE

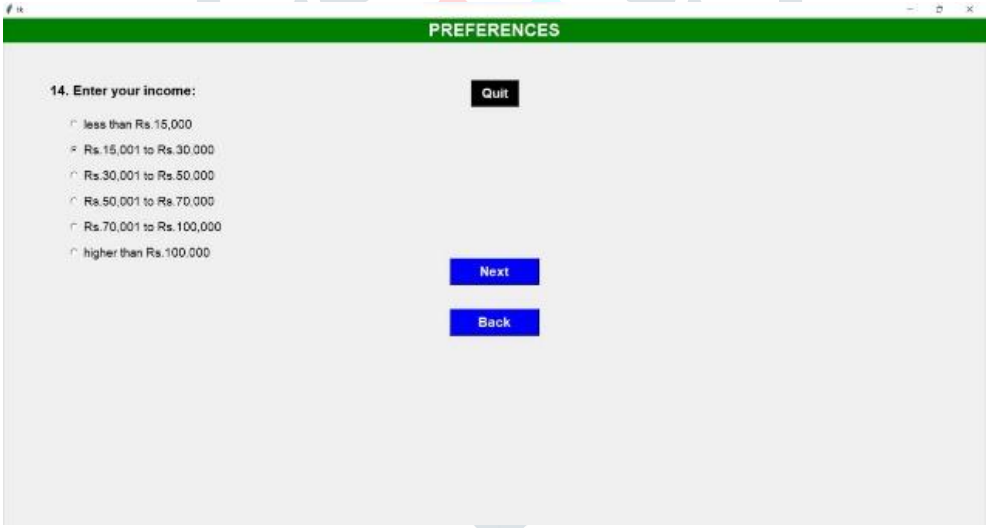


Figure -IV

LOCATION SELECTION

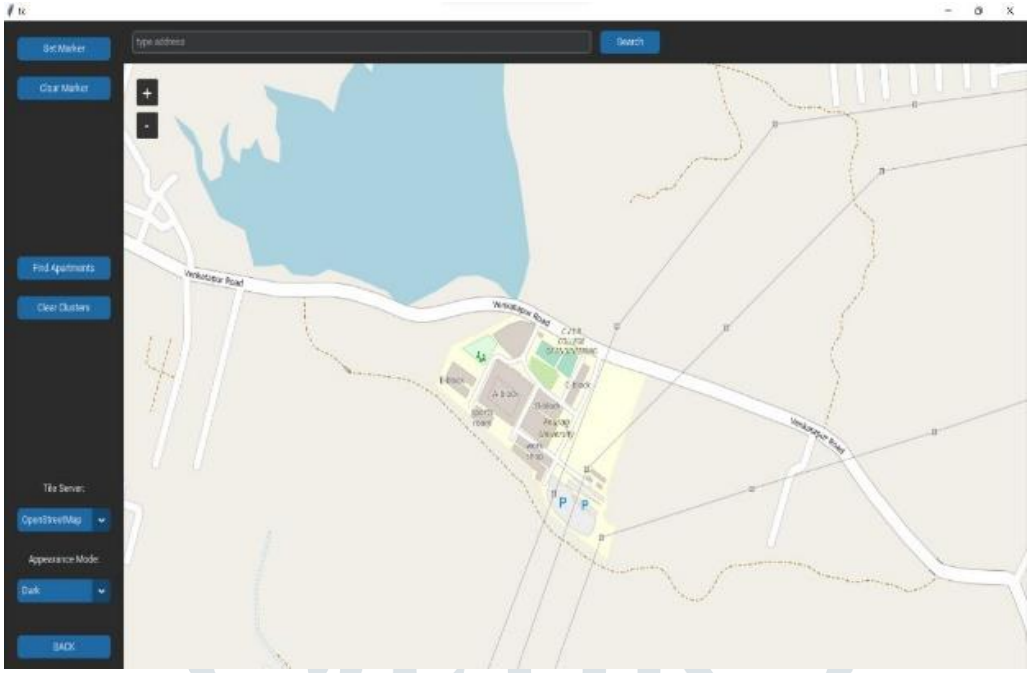


Figure -V

PROCESSING

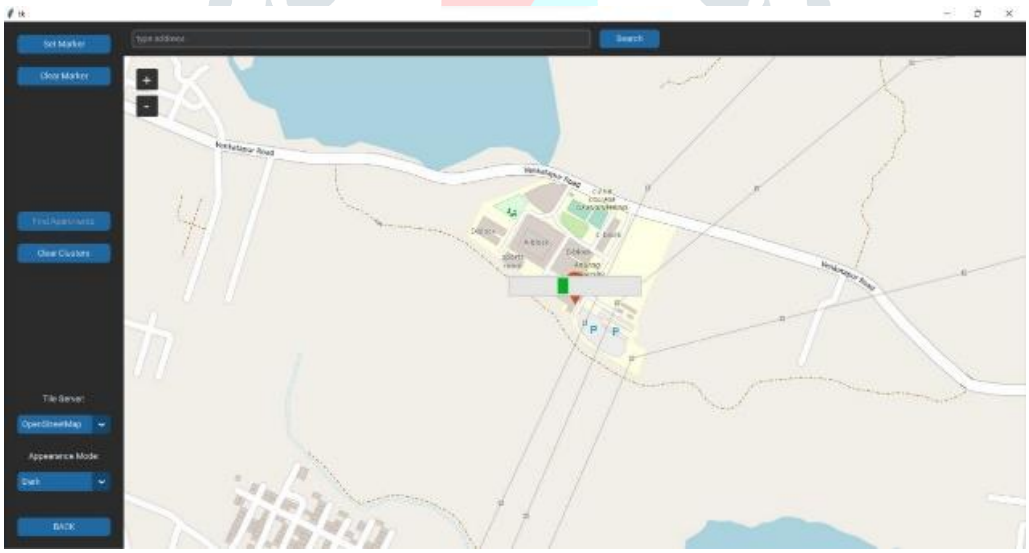


Figure -VI

BEST SUGGESTED ACCOMMODIATION

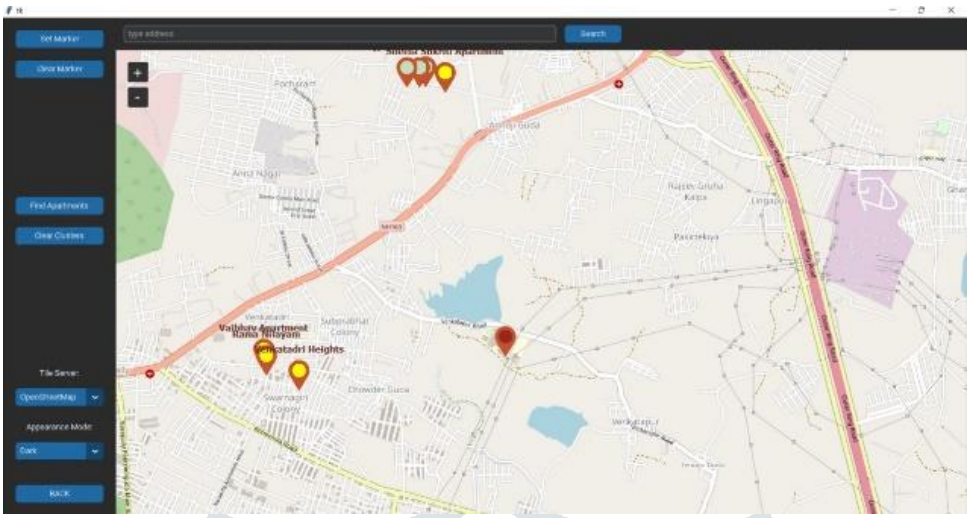


Figure -VII

VIII. CONCLUSION

This research paper outlines the presented methodology for accommodation recommendation based on user input parameters using artificial neural networks. The presented system accepts user requirements as well as a dataset, which is in the form of an excel file that is first preprocessed. Preprocessing entails removing redundant and incomplete data before feeding it into the system. The preprocessed dataset and user input parameters are clustered using the k nearest neighbour clustering mechanism. The clustering is done with the user input parameters, and the results are sent to the next step for neuron generation.The if-then rules are used in the decision making approach to classify the probability list. The resulting output is displayed to the user via the interactive user interface. Based on the user's preferences, the best top three accommodation suggestions are displayed.

IX. ACKNOWLEDGEMENT

The study of geolocation data is a complicated, multi-disciplinary field of research that has been made feasible by the efforts of numerous academics and organisations. As an AI language model, I would want to thank these people and organisations for their significant contributions, without which this field of study would not be conceivable.

X. REFERENCES

1. Kriegel, Hans-Peter; Schubert, Erich; Zimek, Arthur (2016). "The (black) art of runtime evaluation: Are we comparing algorithms or implementations?". Knowledge and Information Systems. 52 (2): 341–378. doi:10.1007/s10115-016-1004-2. ISSN 0219-1377. S2CID 40772241.
2. L. Nikdel, A.E.S. Schay, D. Hou, S.E. Powers.Data-driven occupancy profiles for apartment-style student housingEnergy Build., 246 (Sep. 2021), Article 111070, 10.1016/j.enbuild.2021.111070

3. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron (2019) is a more advanced book that covers machine learning using Python.
4. Yang, C., Yang, L., Li, C., Chen, Y., & Li, F. (2018). Mining geolocation data for disease surveillance: a review. International Journal of Environmental Research and Public Health, 15(8), 1670.
5. Islam, S., Hassan, M. M., & Almogren, A. (2020). Exploring geolocation data for smart mobility applications: a review. IEEE Access, 8, 161764-161782.
6. Islam, M. T., Hasan, M. A., & Choudhury, N. (2019). Exploring geolocation data for social media analytics: a review. Journal of Big Data, 6(1), 58.
7. Li, C., Yang, C., & Zheng, X. (2019). Exploring geolocation data for tourism and hospitality: a review. Journal of Hospitality and Tourism Technology, 10(4), 496-511.
8. Hussain, M., Ali, S., Shah, S. A. A., & Ahmad, I. (2021). Exploring geolocation data for crime analysis: a review. Journal of Ambient Intelligence and Humanized Computing, 12(4), 3437-3454.
9. M. Sumithra and S. Malathi, "A survey on Medical Image Segmentation Methods with Different Modalitites",
10. International Journal of Engineering Research and Technology (IJERT) – Vol. 6 Issue 2, Mar 2018.
11. B.Buvaneswari and Dr.T. KalpalathaReddy,"ELSA- A Novel Technique to Predict Parkinson's Disease in BioFacial",International Journal of Advanced Trends in Computer Science and Engineering, ISSN 2278-3091,Vol.8,No.1,Pp. 12-17,2019