IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124109

Exploratory Geolocational Data

Aamna Malik[1], Khushbu Shah[2], Sofiya Sheikh[3], Aman Wagh[4], Samina Anjum[5]

Student, Computer Science and Engineering, Anjuman College of Engineering and Technology, Nagpur, India[1-4]

Assistant Professor, Computer Science and Engineering, Anjuman College of Engineering and Technology, Nagpur, India[5]

Abstract: This project involves the use of K-Means Clustering to find the best accommodation for the students in Nagpur by classifying accommodation for incoming students on the basis of their preferences on amenity, budget and proximity to the location.

Keywords: Data, K-Means, Recommendations, Map

I. INTRODUCTION

In the fast-paced and busy environment where the average person lives, it is common for people to be too tired to prepare home-cooked meals. And of course, even if you eat homemade food every day, it's not unusual if you want to go out to eat once in a while for social/recreational purposes. However, it is a commonly understood idea that no matter where you live, the food you eat is an important aspect of the lifestyle you live. Now, imagine a scenario where someone has just moved to a new location. They already have certain preferences, certain tastes. This would save students and food suppliers a lot of trouble if the student lives near his or her favorite outlet.

II. RELATED WORK

A. Fetch Datasets from the relevant locations

If you want to do data analysis, you need to get the dataset and you need to setup the environment required for the data analysis.

B. Clean the Datasets to prepare them for analysis

After getting the data and understand what it says. A best way to do this is by visualising the data via graphs. Graphs help us quickly get an easy understanding of the data, and are a much more User-friendly.

C. Visualise the data using box plots

K Means Clustering will help us cluster the places based on the facilities located around them. A location with more facilities nearby will be said as “Amenity Rich" while a location with less facility as "Amenity Poor". Similar locations will be grouped. Run the K-Means Algorithm and find out the best value for K, which we will use in our app.

D. Fetch Geolocational Data from the Foursquare API

After getting better cluster values, you need to get Geolocational data from the Foursquare API to find some accommodation for our users.

E. Plot the Cluster locations on the Map

At last plot the results on the map that will the user to view the location.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124109

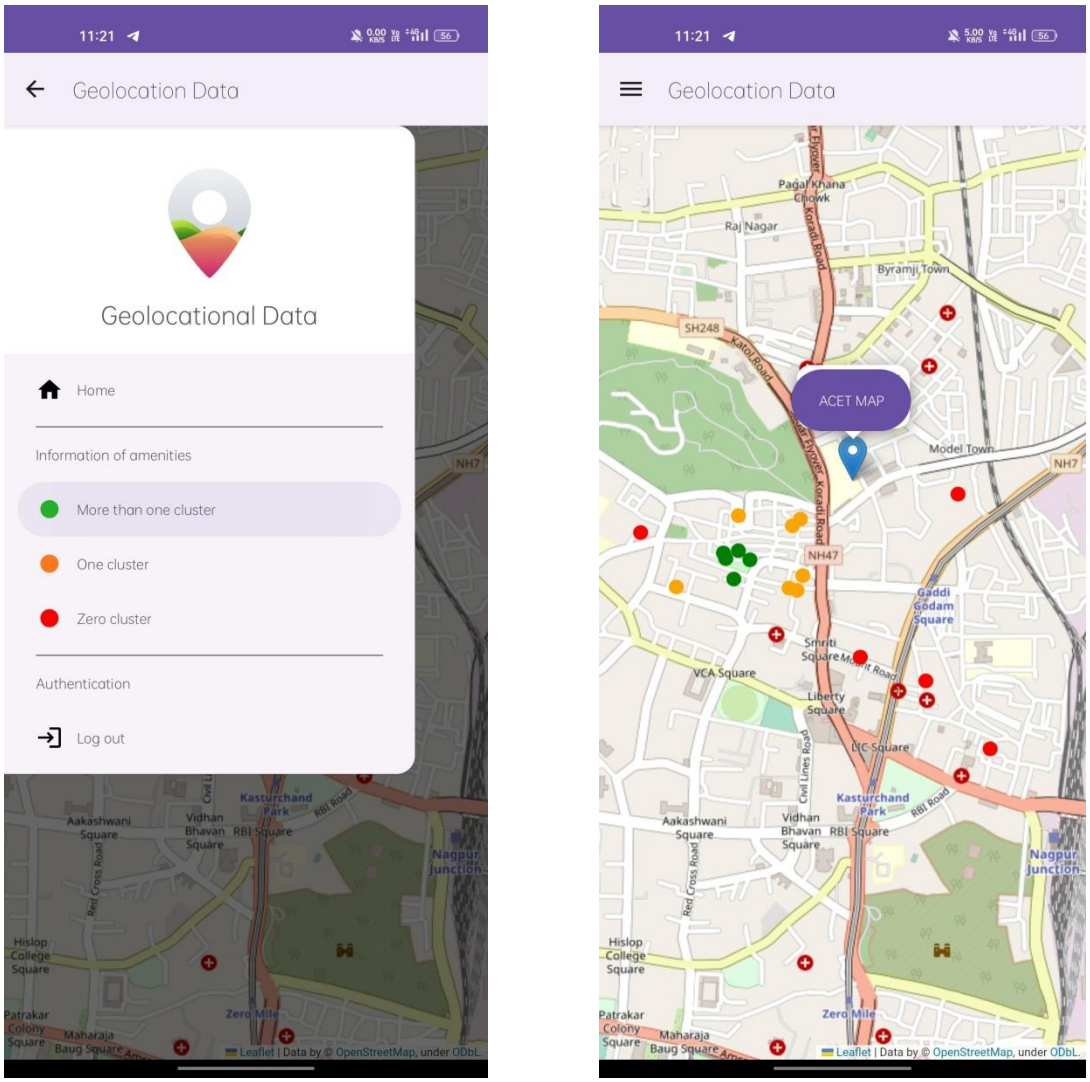

FIGURE 1: Cluster on the Map

III. EXISTING SYSTEM

The existing system contains hostels and apartments for rent and also it has buy and sell options. It doesn't recommend accommodation in our budget. It has rare cases of houses on our budget. It also doesn't recommend restaurants, gyms etc., based on users' budget. Previous researches lack accuracy of true recommendations.

IV. PROPOSED SYSTEM

Exploratory Geolocational data analysis involves analyzing data related to location and spatial relationships. Here's a proposed system for conducting exploratory Geolocational data analysis. The first step is to define the research questions that you want to answer through Geolocational data analysis. This may include questions related to identifying patterns in location-based data, understanding spatial relationships between different variables, and predicting future trends based on historical data. Then need to collect and preprocess the Geolocational data that will be used for analysis. This may involve gathering data from various sources such as GPS devices, social media platforms, and public data repositories. The data should be cleaned and processed to ensure accuracy and consistency. Once the data has been collected and preprocessed, it should be visualized using appropriate geospatial tools and techniques. This may include creating maps, heat maps, and other visualizations that help to identify patterns and trends in the data. The next step is to conduct exploratory data analysis to identify spatial relationships between different variables. This may involve using statistical techniques such as correlation analysis, regression analysis, and clustering to identify patterns in the data. Finally, the results of the analysis should be interpreted to draw insights and conclusions that can inform

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124109

decision-making. This may involve identifying areas for further investigation, developing predictive models based on historical data, and making recommendations for future actions based on the analysis. Overall, an effective exploratory Geolocational data analysis system should be designed to gather, preprocess, visualize, analyze, and interpret location-based data in a way that allows researchers and decision-makers to identify patterns, relationships, and trends in the data that can be used to inform decision-making.

A. Way of Approach

- Get Datasets from the relevant locations (Data Collection)
- Clean the Datasets to prepare them for analysis. (Data Cleaning via Pandas)
- Visualise the data using boxplots. (Using Matplotlib /Pandas)
- Fetch Geolocational Data from the Foursquare API. (REST APIs)
- Use K-Means Clustering algorithm to cluster the locations (Using ScikitLearn)
- Present findings on the map. (Using Folium/Seaborn)

B. Project stages

FIGURE 2: Data Flow Diagram

C. Modules

1. Data Collection Module
2. Searching Module
3. Suggestion Module
4. Communication Module

4.C.1 Data Collection Module
Collect the data from the users and store the data in the database for later use.

4.C.2 Searching Module
After giving the input the user search for the location comes under their budget with their required facilities.

4.C.3 Recommendation Module
After searching the required information on the search bar, it will show the recommendation. Based on budget and requirements.

4.C.4 Communication Module
From the shown recommendations, the user selects the best accommodation and with the help of communication module it will redirect the user to the contact details of the owner.

V. CONCLUSION

The literature survey helped us understand how machine learning is majorly used for geostatistics and spatial application. Also by comparing the pros and cons of different techniques used by recommendation systems, it was concluded that knowledge-based filtering is best suited for housing recommendation systems. To build this system an environment was set up using Jupiter Lab. Pandas library was used for data cleaning. Box plots helped in providing a visual representation of data for better analysis. K-Means Clustering was used for model training. This ML model was connected to Foursquare API from which residential locations in a fixed radius along with basic amenities our extracted using an HTTP library called requests. Lastly using Folium the results is presented on a map. This app is easy to use because it is user-friendly and budget-friendly. A common problem of the migrant people is solved through this app. This app is used to find accommodation easily and its fits in your budget and it will be more useful for students who are studying.

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124109

REFERENCES

[1]. M. Sumithra and Dr. S. Malathi, "A Novel Distributed Matching Global and Local Fuzzy Clustering (DMGLFC) FOR 3D Brain Image Segmentation for Tumour Detection", IETE Journal of Research, doi.org/10.1080/03772063.2022.2027284, 2021

[2].B.Buvanswari and T.Kalpalatha Reddy, "A Review of EEG Based Human Facial Expression Recognition Systems in Cognitive Sciences" International Conference on Enenrgy, Communication,Data analytics and SoftComputing(ICECDS),CFP17M55-PRJ:978-1-5386-1886-8",August 2017.

[3] .Cethana, C., Subbiah Swaminathan, S. Sharanyaa, E. Sathish, R. Prathipa, and Anuradha Thakare. "Application Of Reverse Engineering in the Process of Utilization of Human Brain in Artificial Intelligence." Journal of Optoelectronics Laser 41, no. 3 (2022): 89-93.

[4] .M. Sumithra and Dr. S. Malathi," Modified Global Floyyou Pollination Algorithm-based image fusion for medical diagnosis using computed tomography and magnetic resonance imaging", International Journal of Imaging Systems and Technology, Vol. 31, Issue No.1, pp. 223-235, 2021

[5]. K. Sridharan, and Dr. M. Chitra"SBPE: A paradigm Approach for proficient Information Retrieval, Jekyll Journal, and "Vol 63, No. 7 Jul 2013

[6]. Sharanyaa, S., P. N. Renjith, and K. Ramesh. "Classification of Parkinson's disease using speech attributes with parametric and nonparametric machine learning techniques." 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS). IEEE, 2020.

[7] M. Sumithra and Dr. S. Malathi, "3D Densealex NET Model with Back Propagation for Brain Tumour Segmentation", International Journal of Current Research and Review, Vol. 13, Issue 12, 2021.

[8].B.Buvanswari and T.Kalpalatha Reddy, "EEG signal classification using soft computing techniques for brain disease diagnosis", Journal of International Pharmaceutical Research ,ISSN : 1674-0440,Vol.46,No.1,Pp.525-528,2019.

[9] Sharanyaa, S., P. N. Renjith, and K. Ramesh. "An Exploration on Feature Extraction and Classification Techniques for Dysphonic Speech Disorder in Parkinson's Disease." In Inventive Communication and Computational Technologies, pp. 33- 48. Springer, Singapore, 2022.

