Geolocation Analysis System Using K-Mean Clustering

Siddhi Mohite[1], Pruthviraj Aher[2], Prof. Amit Kukreja[3]

[1,2,3]Department of Electronics and Telecommunication Engineering K.J. Somaiya Institute of Technology, Mumbai, India

ABSTRACT:

In recent years, immigration has increased, particularly among students and employees seeking education or work opportunities. However, moving to a new place brings various challenges, with accommodation being a significant concern. Students and young adults face difficulties finding suitable housing that meets their preferences and budget [1].

To address this, we propose using K-Means Clustering, a data analysis technique, to recommend accommodations for migrants based on factors like facilities, budget, and proximity to their destination. By analyzing geo-locational data, we can classify different types of accommodations such as PGs, flats, hostels, and nearby amenities like food stalls and daily necessities [1].

The main goal is to develop a system that suggests the best accommodation options near KJSIT College, tailored to individual preferences. This system can be expanded for various purposes, such as helping businesses locate themselves strategically near educational institutions for better accessibility to potential customers.

INTRODUCTION:

In today's digital age, the internet offers a vast repository of information, shaping our decisions in various aspects of life. From what we eat to where we stay, the internet plays a pivotal role, largely through recommendation systems that tailor suggestions based on our preferences and online behavior [3].

Migration, particularly in India, is on the rise, driven by factors like employment and education opportunities. However, relocating to unfamiliar places, especially for students, poses challenges such as finding suitable accommodation without local knowledge or connections [1].

To address this issue, a project is underway specifically for students of KJSIT College, aiming to simplify the process of finding accommodation near the campus. This project employs algorithms like K-Means clustering and content-based filtering to recommend hostels, paying guest rooms, and other facilities based on the student's preferences and location [1].

In essence, the study focuses on providing an efficient solution for student accommodation needs, leveraging technology to streamline the search process and enhance the overall experience of transitioning to a new environment for education.

LITERATURE SURVEY:

Recommendation systems have become essential in the digital age, helping users find what they need amidst the vast sea of information online. By analyzing user data, these systems offer personalized

suggestions for movies, books, music, and more [3]. For instance, if you enjoyed a certain movie, the system might suggest similar ones you might like. These recommendations streamline decision-making and reduce information overload [3].

Recently, there's been a trend toward incorporating location data into recommendation systems. This means recommendations can be even more relevant, taking into account where you are. Academic and social network fields are taking notice of this trend, recognizing its potential to improve user experiences.

One powerful technique for recommendation systems is K-means clustering. This method groups similar data together, making it easier to identify patterns and make recommendations. However, noisy data and outliers can sometimes affect the accuracy of these recommendations [1].

To address this, there's a new approach that gives more weight to reviewers whose tastes align closely with the user's preferences [1]. This helps tailor recommendations even further, increasing user satisfaction. Testing has shown promising results, especially in suggesting hotels based on user preferences.

PROBLEM STATEMENT:

Imagine someone, like a staff member or student, who has just moved to a new area. Finding the perfect place to stay can be tricky because they have specific preferences regarding what they want in terms of amenities, cost, and location [1].

Our project, the Geolocation Analysis System, aims to solve this problem by recommending the best lodging option for the user based on their preferences. Additionally, it provides information about local transportation options that can help them reach their destination easily.

OBJECTIVES:

In our project, geolocation analysis system, our objectives are as follows:

1. **Retrieve Available Accommodations**: Our goal is to compile data on every hostel and paying guest (PG) room that is available in the user's city. This entails gathering in-depth information on different lodging choices.

2. **Filtering Based on User Requirements:** After compiling the options, we plan to filter these PG and hostel rooms according to the user's specified preferences. This could involve elements like your ideal location, preferred facilities, budget, etc.

3. **Provide Transportation Options:** Our goal is to provide consumers with information regarding the closest transit options from the chosen dorm or guest room. This could entail displaying options like bus schedules, metro stations, or ride-sharing services by integrating with transportation APIs.

4. **Implement Recommendation System for Affordability:** Our system will make recommendations to users for reasonably priced PG and hostel rooms using a recommendation engine. This could depend on things like the user's budget, past pricing information, and availability.

5. **Estimate Price Based on User Preferences:** Lastly, depending on customer preferences, we try to give them an approximate price range for PG and hostel rooms. This can entail looking at pricing information gathered from multiple sources and taking demand, facilities, and location into account.

Our system's overall goal is to make it easier for immigrants to find acceptable housing by giving them access to detailed information, allowing them to select possibilities based on their preferences, and making customized recommendations.

METHODOLOGY

We are developing a project to gather and arrange data on various cities. To start, we'll purge the data of any mistakes or odd values. Next, we'll use a technique known as K-means to group together areas that are comparable. We will collect information on paid guest rooms and hostels in each area with the use of the Foursquare API. We'll match these alternatives with user preferences using a method known as content-based filtering. Lastly, we'll use folium or other mapping tools to plot all of this data on a map.

The steps for our project are [2] :
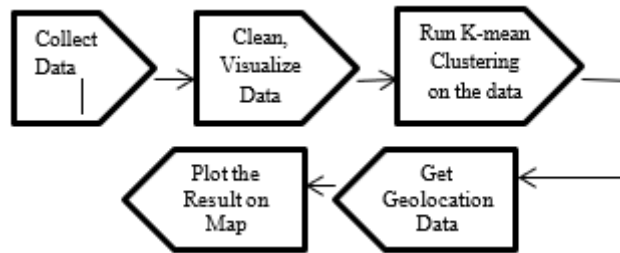
Fig. Block Diagram

1. **Collect Data:** Get the dataset from the relevant location. preparing the environment for data analysis. selecting from a plethora of possibilities the relevant and essential qualities (money, commute to job or college, etc.).

2. **Clean, Visualize Data:** Cleaning and getting ready for analysis of the dataset. The survey data has several abnormalities that could lead to unexpected and erroneous results, making it raw and unsuitable for analysis.

3. **Run K-Mean Clustering:** The data point is sent to the cluster that is closest to the cluster center. Once more, the new cluster is computed. Once the data has been gathered, graphs are used to evaluate and illustrate the results. A useful tool for comparing distributions across several groups and datasets is the general analysis of trends.

4. **Get Geolocation Data:** To group the data, K-means clustering is being applied. K-Means clustering is applied to the population dataset, separating the data into discrete groups based on shared features. It is one of the simplest unsupervised learning strategies to train the machine learning model when dealing with unlabeled input. It distributes data points to clusters so as to guarantee that there is a minimum sum of squared distance between data points and the cluster center. As variation decreases, so does the similarity between data points.

5. **Plot the Result on Map:** The location is clustered using K-means clustering. The best site to stay is found by using K-means clustering to the final location data; the results are then plotted on a map using Map.

RESULT:

K-means clustering is used to cluster the location. Using K-means clustering to the final location data, the optimal place to stay is determined; the findings are then projected on a map using Map.
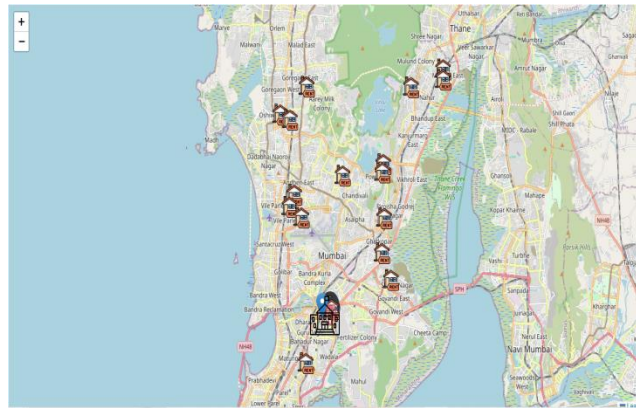
.Based on geography and user preferences, the model recommends hostels that are in the same cluster as the user's input. When recommending content, ratings and distance from the user's location are taken into consideration.

The model output provides a list of recommended hostels together with their names, addresses, prices, ratings, and distances from the user's present position. Thanks to the information offered, customers may choose a hostel with confidence based on their preferences and location [1].

CONCLUSION:

Our technology finds reasonably priced guest rooms or hostels based on your location and choices. K-Means Clustering is being used to determine which local lodging options are most suitable for KJSIT students. Finding appropriate housing is a typical issue that our product addresses, especially for students on a tight budget. It is also user-friendly.

REFERENCES:

1. Basavesh D1, Laharishree S2, Sthuthi S3, Tejaswini N4, Vidya R5, "Hostel Finder: Location-Based Recommendation System for Hostels and PGS with Transit Information". International Journal of Research Publication and Reviews, Vol 4, no 5, pp. 4482-4489 May 2023.
2. K. Gayathri, M. Vijayalakshmi , M. Kousalya, N. Manikiran , "Geolocation Data Analysis By Applying K-Means Clustering". International Journal of Science Technology and Management, Vol.12, Issue No.04,2023.
3. Sakshi Rajesh Sinha, Prof. Sumedh Pundkar, "Geolocation Analysis Using Machine Learning". International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 9, Issue 6, June 2022.
4. Hui Xu, Shunyu Yao, Qianyun Li, Zhiwei Ye. An Improved K-means Clustering Algorithm. IEEE International Symposium September 2020.
5. O. Aboulola, "A literature review of spatial location analysis for retail site selection," Journal of the Association for Information Systems, 08 2017.
6. Jun, H. Jong, Kim, J. Hee, Rhee, D. Young, Chang, and Woo, ""seoulhouse2vec": An embedding-based collaborative filtering housing recommender system for analyzing housing preference," Sustainability, vol. 2, no. 7, 2020. [Online]. Available: https://www.mdpi.com/20711050/12/17/6964
7. Jitechana, Lubna Shaikh, Shefali Bhattacharjee, Vaishali Yenolge, Dr. N. P. Kulkarni. An Accommodation Recommendation System for Immigrants using Exploratory Data Analysis on Geolocation Data. IJARSCT 2022.