The Battle of Neighborhoods

Shivam Goswami

25 July 2020

A. Introduction

A.1 Description and discussion of background

This project aims to find similar neighborhoods in city of Delhi, India for an individual moving from Dabri Village, Dwarka to Connaught Place, New Delhi due to better economic opportunities.

Our subject has his house located in Dwarka, he loves his neighborhood mainly because of all the amenities and facilities such as parks, pharmacies, schools, malls, shopping centers, hospitals he gets in his neighborhood.

He receives a very good job offer from a reputed company located on the opposite side of the city i.e. New Delhi. However, given the far distance from his current place if he decides to take up the job offer, he must relocate to the New Delhi. He is willing to take up the job offer but wants to move to a neighborhood in New Delhi which is like his current neighborhood in Dwarka.

In this project we will help our subject find a similar neighborhood in New Delhi.

A.2 Data Description

We will be dealing with the following data in our project:-

- Dataset containing Borough and Neighborhood of Delhi, the dataset contains the names of Borough and Neighborhood in Delhi. Here is a link to dataset dataset .
- I used python library **geopy** to extract coordinates of neighborhoods and saved them into a csv file on my computer. Data required some scraping for **geopy** to correctly give the coordinates of boroughs of Dwarka and New Delhi, since they are our points of concern. Here is a link to dataset coordinates_dataset .
- I used Foursquare API to explore the most common places of a neighborhood in form of a json file.

B. Methodology

As dataset, i used coordinates_dataset , which i prepared using Election commission datset. This dataset contains Borough, Neighborhood, Latitude and Longitude of the location.

I used python library **folium** to visualize Delhi, its boroughs and neighborhoods.

I used Foursquare API to explore the neighborhoods and used json file returned by Foursquare API to extract names and categories of venues nearby our neighborhoods in Dwarka and New Delhi.

The data also showed that there were 125 unique venue categories in all neighborhoods of New Delhi and Neighborhood of Dabri Village.

Since, our subject needs to find a neighborhood like Dabri village, Dwarka in New Delhi area, i dropped all the other irrelevant neighborhoods of Dwarka. Now our dataset only contains the parameters of Dabri village and neighborhoods of New Delhi.

I then used machine learning algorithm **K means** to cluster the neighborhoods of New Delhi and Dabri village, Dwarka.

K Means is a non parametric clustering algorithm used for clustering data. K-means clustering is an unsupervised clustering algorithm that gathers and groups data into k number of clusters. K-means divides the data into non-overlapping subsets (clusters) without any cluster-internal structure. The data within a cluster is very similar, data across different clusters is very different. The variable k is considered to be a parameter that will be established by the machine learning engineer. I chose k=5 after running accuracy tests on the data.

Finally I used python library **geopy** to compute distances between coordinates taking into account the elliptical curvature of Earth.

C. Result

After doing the fit on data, the model returned 5 clusters, namely cluster 0, cluster 1, cluster 2, cluster 3, cluster 4. each cluster comprises of many neighborhood which are similar to each other. This means neighborhoods or localities which are in cluster 0 are similar to all other neighborhoods or localities in cluster 0. same goes with other clusters. The algorithm computes the similarity between neighborhood by computing the Euclidian distance, the lesser the Euclidian distance, more the similarity.

There were total 29 neighborhoods in cluster 0 excluding Dabri village.

Since, there are 29 neighborhoods, where should our subject should relocate to. His first criterion of similarity is fulfilled. I think he will now focus on the distance between these neighborhoods and Connaught place, remember he got a job offer from a reputed company located in Connaught place.

I calculated the distance between these neighborhoods and Connaught place and found that Peshwa Road locality in New Delhi is nearest to Connaught place and this locality is very similar to Dabri village.

D. Discussion

Delhi is a big city with diverse boroughs and neighborhoods. Different clustering algorithms can yield different clustering.

I used K Means technique, since I have good past experience with it. I tested the algorithm for different values of k ranging from 1 to 9 and checked the accuracy scores with each value of k. I found k=5 to be the optimum value of k. The coordinates of all the neighborhoods of New Delhi and Dabri village were used. Some neighborhoods didn't return popular venues, hence, their cluster couldn't be determined by K means algorithm, although their number is very small and hence they do not deviate our accuracy.

Python library **folium** has been used to create map visualizations of cluster data.

I ended study by finding the nearest neighborhood from Connaught place in New Delhi. In future studies some other criteria can be carried out.

F. Conclusion

Relocation due to better economic opportunities is very common now a days. Studies like this can help people better adjust to their new neighborhood.

Real Estate companies can help their customers by providing better and smooth relocation by using similar data analysis.

Thank you

Shivam Goswami