INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

Cluster Algorithm Based Neighbourhood Exploration

Bharanidharan A[1], Kayalvizhi V[2], Lavan Harshaini H[3]

Assistant Professor, Dept. of Computer Science and Engineering, Sri Ramakrishna Engineering College, Coimbatore, India[1]

UG Student, Dept. of Computer Science and Engineering, Sri Ramakrishna Engineering College, Coimbatore, India[2,3]

ABSTRACT: This paper is designed to create a cluster of similar restaurants which represents the best places and where it is located. Not everyone has the same tastes and likes, so when a person asks for a recommendation or a suggestion of a best restaurant in the neighbourhood, one usually gets a biased recommendation of limited number of restaurants, which in case may not be the best one for the enquirer. To get the best recommendation, it is always the best to trust the publics opinion and compare among top rated places to hangout in city, in this case its New Delhi. The purpose of the study is to make it easier for the tourists and localities to identify some of the coolest places. In this study 5 clusters were built, where each cluster represent which is best and where it is located. It can recommend the best for people who want to explore and enjoy a city like New Delhi.

KEYWORDS: Best Restaurants, Clustering, Foursquare API, New Delhi, Exploring Neighbourhoods', Tourism, Machine Learning, Zomato.

I. INTRODUCTION

New Delhi is the capital city of India which attracts a greater number of tourists. It is the part of the town of Delhi's eleven districts. The city itself is a populace that exceeds 26 million. New Delhi is used interchangeably to consult the National Capital Territory of Delhi (NCT). The Official Language of New Delhi and the one this is most broadly spoken in Hindi. Nevertheless, English is spoken as a language within corporations and many agencies. Over the final decades, it is continuously growing due to the metropolis's vital role in authorities and business. With its diverse culture, there is an diversity in the variety of food provided in New Delhi. Lot of restaurants are available which cater to different taste of the New Delhi population. They vary in cuisines, décor and ambience. Most of them can be categorized based on their cuisines, décor, service and ambience. Most of them can be categorized based on their cuisines such as Italian, Chinese, French, etc,. Out of all these places, it would be difficult for a Tourist to choose a place which meets the need of the person. The tourists get to choose only from a small number of categories based on others recommendation which may not be the best one for the receiver. Hence a group of all the places are grouped into clusters based on their categories which makes it easier to compare between the placed based on the requirements.

II.RELATED WORK

Zomato - Market and consumer analysis:
Aniruddha Deshpande has said that a few start-ups in the Indian Food & Foodstuffs enterprise have transformed the way we go searching for places to dine. The scenario has changed from previous exercise while we had to visit restaurants to choose the eating place and have the experience. Today, the choice for the first-rate locations to have food is just a click on away. One can pick out the fine-rated region and then decide to experience it there with friends and family.

The purpose of this case has a look at is to do an analysis of the way Zomato has scaled up its operations, expanded its enterprise into numerous countries. We additionally plan to find out the strengths, weaknesses, future opportunities, and threats of their enterprise model. This will in the long run help us recognize the F&B enterprise and construct our evaluations approximately few commercial enterprise scenarios so that it will assist us build answers to them. Questions that can be asked using the above-mentioned problem.

1. What is the best location in New Delhi city for Chinese Cuisine?
2. What is the best location in New Delhi city for Cafeteria and Pizza?
3. Which areas have a large number of Chinese restaurant market?

4. Which location has the best night club in New Delhi city?
5. Which is the best tourist place in New Delhi?
6. Which areas has less number of restaurant?
7. Which is the best place to stay if I prefer Chinese restaurant?
8. What areas have the best fast-food items?

II. PROPOSED ALGORITHM

A. *Collection of data:*

Thedatasetsarecollectedfromkaggle.The datasets have used for creating the clusters of places based on their categories which consists of best places of New Delhi.

    1. Getting all the localities of New Delhi:

        By using Four square API we will get all the venues in each neighbourhood.

    2. Exploratory Analysis:

- Collect the New Delhi city data from Kaggle.
- Using Foursquare API, we will find all the venues for each neighbourhood.
- Filter out all venues that are nearby locality.
- Using the aggregative for each restaurant to find the best places.
- Visualize ranking of neighbourhoods using folium library (python).

B. *Machine Learning Algorithm:*

Machine learning algorithms is used so that it can make predictions on data sets. It operates by construction of a model to make data-driven predictions or choices rather than following static programming instructions.

C. *Unsupervised Learning:*

The unsupervised learning method is used so that it can fetch out and detect patterns that weren't detected before and it can also deal with unlabelled data and the relationship between them.

    1. K-means Clustering:

        K- means Clustering is a lazy learning algorithm and it classifies a new data point into the target class, counting on the features of its neighbouring data points.

D. *Clustering and Classification*:

Clustering is used to group the data points in the map based on the similarities of the places and are classified based on the similar qualities those data points possess i.e., here it is in case of Indian restaurants, party halls, etc.

    1. Dividing into 5 clusters:

        K-means clustering has been implemented with 5 clusters. The cluster labels are generated for each row in the data frame. New Delhi grouped data is merged with locality venue data to add latitude/longitude for each locality.

    2. Cluster Analysis:There are totally 5 clusters that indicate different restaurants:

- Cluster 0: Red
- Cluster 1: Green
- Cluster 2: Blue
- Cluster 3: Pink
- Cluster 4: Orange.

    3. Creating maps of 5 clusters and grouping into a single Cluster:

- Cluster 0: It is the most recommended for Indian Restaurants.
- Cluster 1: It is the most recommended for Hotels and Nightclubs.
- Cluster 2: It is the most recommended for Fast foods.
- Cluster 3: It is the most recommended for café and pizza shops.
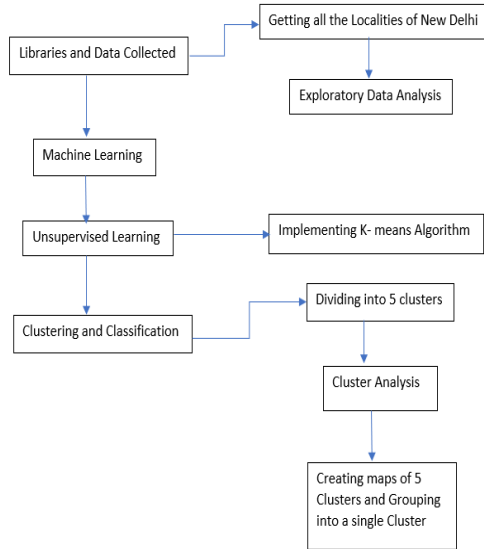- Cluster 4: It is the most recommended for Fast foods.

Fig.1.Flow Diagram

III. PSEUDO CODE

Step 1: Kaggle - Zomato-restaurants-data – is used to fetch the data of New Delhi restaurant and other places data. The data gets fetched by using pandas library. Additionally, the latitude and longitude values for each place are fetched and stored.

Step 2: Using Foursquare API, we will find all venues for each neighbourhood. The Foursquare API offers location data from all over the world for developers. Anyone can create a free developer account to access the location data. Figure below shows Foursquare API credentials.

Step 3: Filter out all venues that are nearby by locality.

Step 4: Grouping is performed using the aggregative rating for each restaurant to find the best places.

Step 5: K-means algorithm is used to create 5 clusters of the locality based on their categories. The cluster labels are generated for each row in the data frame.

Step 6: A final map of all 5 clusters is created by using Folium library (python).

Step 7: End.

IV. SIMULATION RESULTS

I.   Data Analysis:
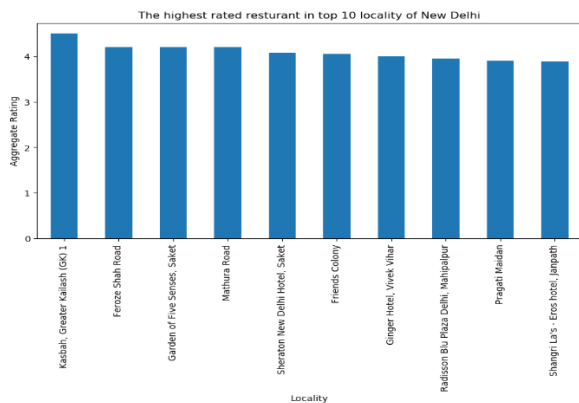
Fig. 2. Localities with Highest ratings

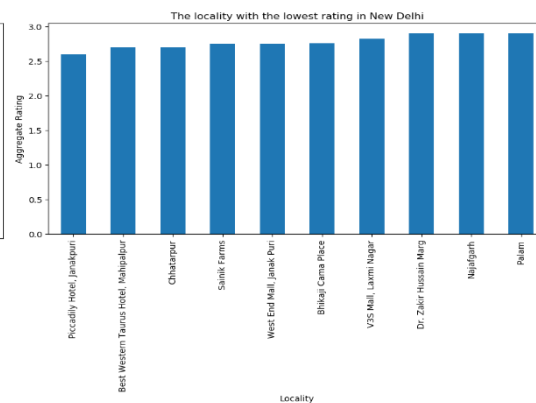The top-rated restaurants are in Kasbah, Greater Kailash area with a rating greater than 4.

Fig. 3. Localities with Lowest ratings.

The worst restaurants are in Piccadily Hotel and Janakpuri with a rating lower than 3.
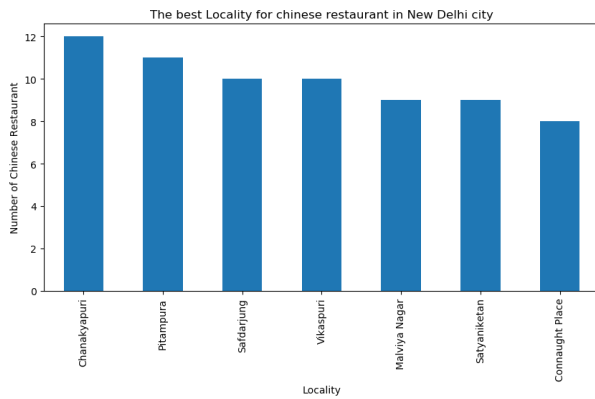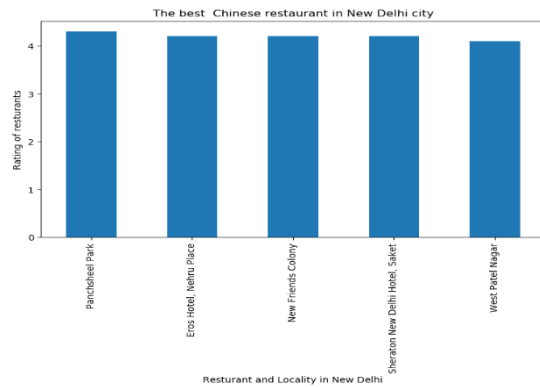
Fig. 4. Best places for Chinese cuisine

Fig. 5. Chinese restaurant with good ratings

Chanakyapuri is the best place for Chinese restaurants in New Delhi. It has almost 12 restaurants.

Panchsheel Park is the best Chinese restaurant in New Delhi and has a rating greater than 4.

II. Final Output:

The Jupyter Notebook is used to implement the cluster and form clusters of places based on their different qualities. The end-user can view places in the map of clusters.
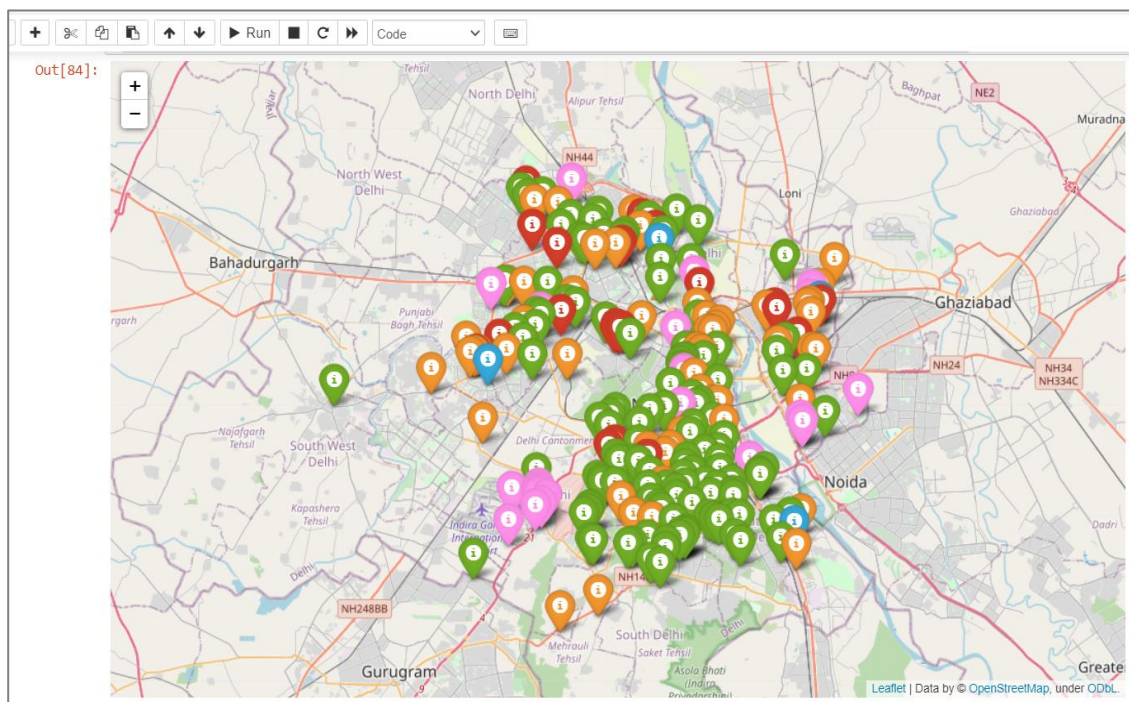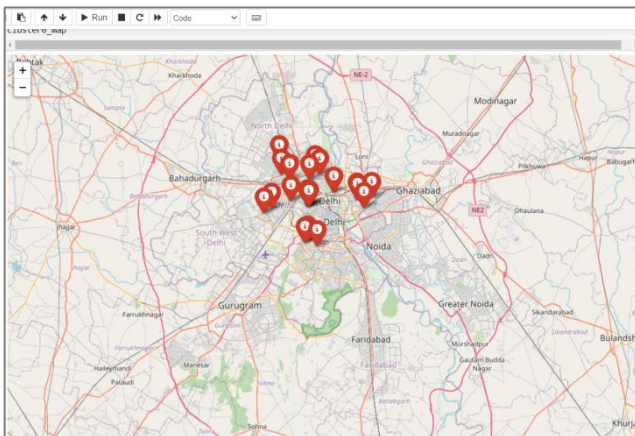

Fig. 2. Final cluster map
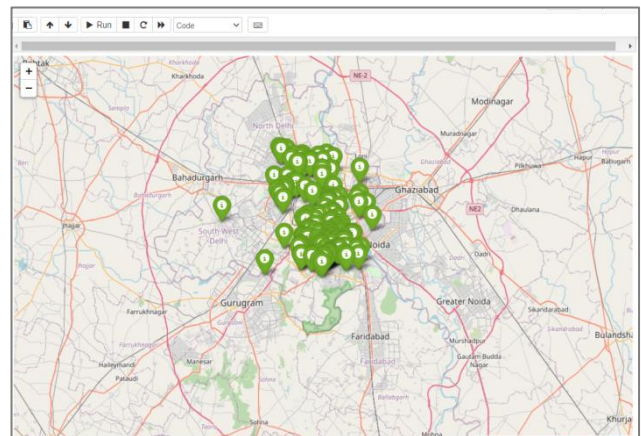
Fig.3.Cluster 0 (Indian Restaurants)

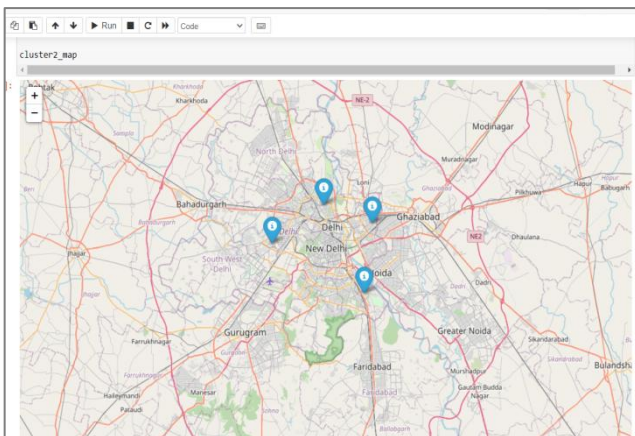Fig.4. Cluster 1 (Hotels and Night clubs)

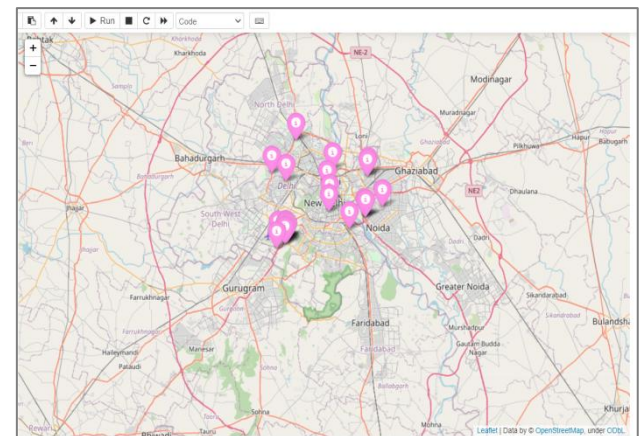Fig.5. Cluster 2 (Fast foods)
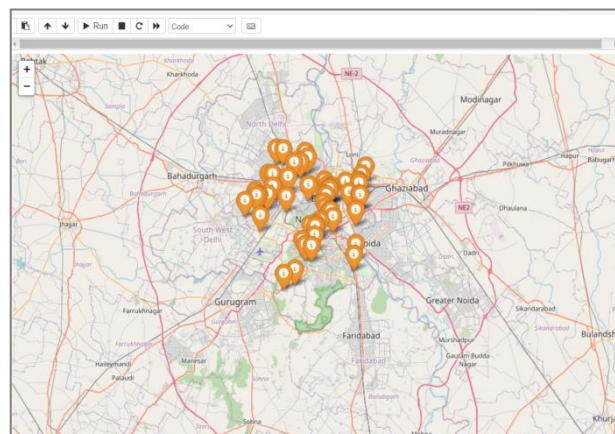

Fig 6. Cluster 3 (Pizza shops and cafeteria)

Fig.7. Cluster 4 (Fast foods)

V. CONCLUSION AND FUTURE WORK

Clustering is an unsupervised method used to group similar data points which is easily understood and manipulated. It is applied in the areas like data mining, pattern recognition, characterizing web document, grouping genes and proteins that have similar functionality, grouping geographical data like earthquakes, etc. We have discussed about various clustering techniques in this paper. From this study we have come to a conclusion that the most suitable method for our project will be k-means clustering.

REFERENCES

1. Cominetti, Matzavinos, Samarasinghe - "Dif FUZZY: A fuzzy clustering algorithm for complex data sets", International Journal of computational intelligence in bioinformatics and system biology, 2010.
2. Anil K. Jain, "Data clustering: 50 years beyond Kmeans", 19th International Conference in Pattern Recognition,2009.
3. Thair Nu Phyu, "Survey of Classification Techniques in Data Mining",IMECS 2009, HongKong.
4. I.K..RavichandraRao,"DataMiningandClusteringTechniques", DRTC Workshop on Semantic Web December 2003.
5. Periklis Andritsos, "Data Clustering Techniques", University of Toronto, 2002.
6. Hung, Yang, "An efficient fuzzy C-means clustering algorithm", IEEE 2001.
7. Amir Ahmad, Lipika Dey, "A k-mean clustering algorithm for mixed numeric and categorical data", Science Direct Data & Knowledge Engineering 63 (2007) 503–527.
8. L. Kaufman, P.J. Rousseew, Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley & Sons, 1990.
9. R. Ng, J. Han, Efficient and effective clustering method for spatial data mining, in: Proceedings of the 20th International Conference on Very Large Data Bases, Santiago, Chile, 1994, pp. 144–155.
10. M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of KDD'96, 1996.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING