

CLUSTER ALGORITHM BASED NEIGHBOURHOOD EXPLORATION TO FIND THE BEST HOTELS IN NEW DELHI

1. SAILESH. S (16RBECSE069)
2. SUJITH. C. SAJEEV (16RBECSE084)
3. SUBHAM SHARMA. A (16RBECSE083)
4. SATHEESH. P (16RBECSE074)

**UNDER THE GUIDANCE OF
DR. B. ARUN KUMAR**

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

**KARPAGAM ACADEMY OF HIGHER EDUCATION,
COIMBATORE.**

ABSTRACT :

Each and every one of us have different tastes and likes, so when a person asks another person for a recommendation of neighbourhood's best restaurants, Hotels, Cafes, Pizzerias, Night Clubs, etc., the enquirer usually gets a biased recommendation which in case may not be the best ones for the enquirer, since his taste may differ. To get the best recommendations, it is always better to trust public's opinion. But finding out which places public likes the best is not easy. This study provides a way for finding out ways for finding best hangouts in a city, in this case New Delhi. The purpose of the study is to make it easier for tourists and localities to identify some of the coolest places. In this study 5 clusters were built, where each cluster represents what 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 :

Data science, Machine learning, Clustering, Four square API, Zomato, New Delhi, Exploring Neighbourhood's, Tourism.

INTRODUCTION :

New Delhi is the capital of India. It is part of the town of Delhi's eleven districts. The city itself has a populace of 257,803. However, the large metro location has 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 is Hindi. Nevertheless, English is also 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. Lots of restaurants are available which cater to different tastes of the New Delhi population. They vary in cuisines, décor, services and ambience.

Most of them can be categorized based on their cuisines such as Chinese, Italian, French, etc.

Here we have categorized the restaurants based on Data Cluster Algorithm:

LITERATURE REVIEW :

AN OVERVIEW OF MACHINE LEARNING AND ITS APPLICATIONS:

Annina Simon et al. has stated that based on their paper appropriate opportunity can be created for the scholars

in their research paper that the opportunity research paper is to create attentiveness amongst upcoming scholars approximately current advances in technology, specially deep getting to know an area of gadget learning which finds applications in massive facts analytics and artificial intelligence.

Deep mastering techniques had been criticized due to the fact there's no way of representing causal relationships (which include among illnesses and their symptoms), and the algorithms fail to collect abstract ideas like “sibling” or “same to.” Not much principle is available for maximum of the methods which can be disadvantageous to beginners.

Data clustering algorithms:

Aman Jatain has stated that this paper is based totally on extraordinary sorts of clustering algorithms with their analogous facts sets has been suggested. In this paper, we've got given a entire relative statistical evaluation of numerous clustering algorithms. Clustering algorithms usually appoint distance metric or similarity metric to cluster the data set into special partitions. Well, regarded clustering algorithms have

been broadly used in numerous disciplines. By seeing the statistical analysis, it is noticed that there may be no highest quality answer for handling issues with big records units of combined and specific attributes. Some of the algorithms can be applied however their performance degrades as the dimensions of facts continues on increasing.

Clustering is the process to find out similarities among any set of statistics. A lot of studies in this clustering subject has emphasised numerical facts units and only a small wide variety of strategies are to be had for specific and different databases.

Indian cuisines: Representing indian culture

Divya Tiwari has said that the scrumptious, fresh, delicious Indian cuisines mirror an 8000-year history of diverse companies and cultures interacting with the subcontinent, main to diversity of flavours and local cuisines found in contemporary India. Indian delicacies consists of a wide sort of traditional and regional cuisines which can be indigenous to India. With the variety of divergence inside the local tradition, geographical location, climate, soil type, etc. those cuisines vary considerably from each other and use locally available spices, herbs, vegetables, etc.

The cuisines replicate the tradition of Indian people and the cause for variety in meals is not handiest the tradition and region however also the availability of spices, staple foods, vegetables, fruits, etc. in the extraordinary areas of the U . S . A .. Indian Cuisine varies from area to vicinity. These cuisines aren't famous best in India however also in overseas countries. The meals displays the lifestyle of India due to the fact we know each pageant of India is tending to entire most effective when a few unique dishes are organized for the festival or unique occasion, like whilst we talk

approximately Eid, our mouth full of water due to the fact with Eid we quickly consider cohesion and Sewai, same as Sewai while we talk about Holi we quickly don't forget colorings and Gujhiya. The cuisine of Indian society suggests the ways of different humans from a exceptional area. It is the only u . S . in which each dish in each u . S . A . may be available.

Recommending popular and personalised venues in a city:

Romain Deveaud has presented Smart Venues, a system that shows nearby venues to a user who visits or lives in a city. Smart Venues models the variant over time of each venue's level of presence and uses today's time series forecasting algorithms to predict the future presence of these venues. We use the predicted tiers of presence to deduce the recognition of a venue at destiny factors in time and to provide the consumer with hints at divergent instances of the day. We also present different interfaces that construct upon our system to show the hints: a internet-primarily based software and a cell application.

We have defined Smart Venues, a device that goals at providing famous and customized venue recommendations to customers, and we've got supplied two interfaces that we constructed on the pinnacle of this system.

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 have 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 Cafe and Pizza?
3. Which areas have a large number of Chinese Restaurant Market?
4. Which location has the best night clubs in New Delhi City?
5. Which is the best tourist place in New Delhi City?
6. Which areas have less number of the restaurant?
7. Which is the best place to stay if I prefer Chinese Cuisine?
8. What places have the best restaurants in New Delhi?

Data source :

1. [kaggle](#) - [Zomato-restaurants-data](#)

- New Delhi Restaurant data that contains list Locality, Restaurant name, Rating along with their latitude and longitude
- Description: This data set contains the required information. And we will use this data set to explore various locality of the new Delhi city.
- Nearby places in each locality of the new Delhi city.

2. Foursquare API.

[Foursquare API](#)

- Description: By using this API we will get all the venues in each neighbourhood.

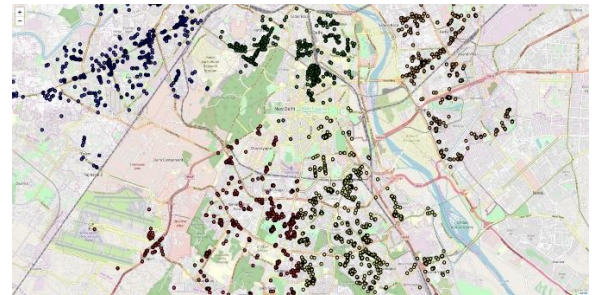
IMPLEMENTATION :

- Collect the new Delhi city data from [kaggle](#)
- Using FourSquare API we will find all venues for each neighbourhood.
- Filter out all venues that are nearby by locality.
- Using the aggregative rating for each restaurant to find the best places.
- Visualize the Ranking of neighbourhoods using the folium library(python).

EXPLORATORY DATA ANALYSIS :

You can find my work at [GitHub Repository](#)

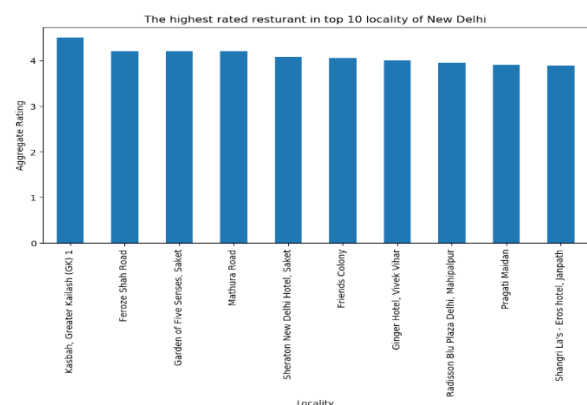
The restaurant clusters:



There are totally 5 clusters that indicate different restaurants

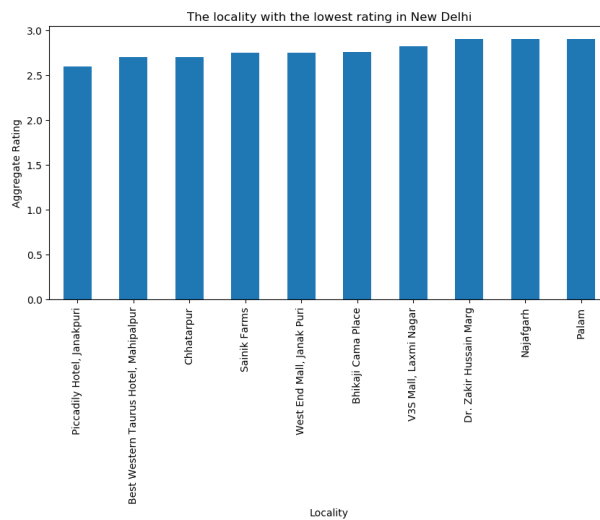
- **Red**
- **Blue**
- **Green**
- **Yellow**
- **Orange**

1. The localities with the highest ratings are:



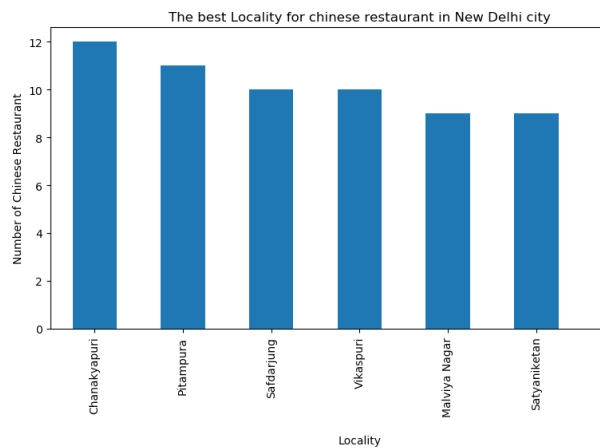
The top-rated restaurants are in Kasbah, Greater Kailash area with a rating greater than 4.

2. The localities with the lowest ratings are:



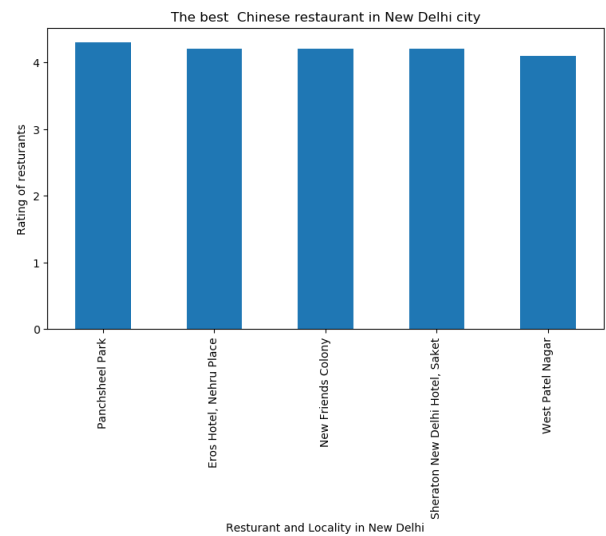
The worst restaurants are in Piccadilly Hotel and Janakpuri with a rating lower than 3.

3. Best places for chinese restaurant in new delhi:



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

4. Chinese restaurants with good rating:



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

GROUPING:

Grouping the data based on the locality:

```
In [45]: # df_Res_Loc = df_Res.groupby('locality').count()[['Restaurant Name']].to_frame()
# df_Res_rating = df_Res.groupby('locality')[['Aggregate rating']].mean().to_frame()
# df_Cuisines = df_Res.groupby('locality')[['cuisines']].agg(' '.join).reset_index()
# df_R = df_Res.groupby('locality')[['Rating (out of 5)']].agg(' '.join).reset_index()
# df_V = df_Res.groupby('locality')[['votes']].sum().to_frame()
# df_Lat = df_Res.groupby('locality').mean()[['latitude']].to_frame()
# df_Long = df_Res.groupby('locality').mean()[['longitude']].to_frame()
# df_Final = pd.merge(df_Lat, df_Long, on='locality').merge(df_Res_Loc, on='locality').merge(df_Res_rating, on='locality').merge(df_Cuisines, on='locality').merge(df_V, on='locality')

In [46]: # df_Final = df_Final[df_Final['Aggregate rating'] >= 4.000000]
# df_Final.columns = ['locality', 'lat', 'long', 'No_of_Restaurant', 'Cuisines', 'Avg_Rating', 'Comments', 'No_of_Votes']

In [47]: # df_Final.head()

Out[47]:
```

	locality	lat	long	No_of_Restaurant	Cuisines	Avg_Rating	Comments	No_of_Votes
0	ARSO Mall Paschim Vihar	28.669345	77.611544	1	North Indian, South Indian, Chinese, Italian, F.	3.100000	Average	117
1	Adhwa	28.537063	77.697800	13	Fast Food, North Indian, Seafood, Continental,	3.282308	Average, Good, Poor, Very Good	1560
2	Aditya Mega Mall, Kirti Nagar	28.656131	77.301286	4	Finger Food, North Indian, Mughlai, Pizza, Fast,	3.275000	Average, Good	434
3	Amrohty	28.933877	77.634270	2	Fast Food, Italian, Pizza, North Indian, South,	3.200000	Average	59
4	Aggarwal City Mall, Pitampura	28.690320	77.134850	3	North Indian, Chinese, Street Food, Italian, Ita.	3.633333	Average	126

The data is grouped by the locality.

FOURSQUARE API:

Defining the Foursquare credentials:

Define Foursquare Credentials and Version

```
In [110]: M CLIENT_ID = 'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx' # your Foursquare .
CLIENT_SECRET = 'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx' # your Foursqu
VERSION = 'yyyyyyyy' # Foursquare API version
```

Getting venues using Four square - api:

All the venues in New Delhi Locality:

```
1. Arss Mall, Paschim Vihar
2. Adchini
3. Aditya Mega Mall, Karkardooma
4. Aero city
5. Aggarwal City Mall, Pitampura
6. Aggarwal City Plaza, Rohini
7. Alaknanda
8. Ambience Mall, Vasant Kunj
9. Anand Lok
10. Anand Vihar
...
238. World mark 1, Aero city
249. Yusuf Sarai
240. Ibis New Delhi, Aero city
There are 211 unique categories.
```

Getting locality along with the top 5 most common venues:

----ARSS Mall, Paschim Vihar----

	venue	freq
0	Indian Restaurant	0.22
1	ATM	0.11
2	Coffee Shop	0.11
3	Pizza Place	0.11
4	Sandwich Place	0.11

----Adchini-----

	venue	freq
0	Café	0.4
1	Pub	0.2
2	Indian Restaurant	0.2
3	Hotel	0.2
4	ATM	0.0

...

----ibis New Delhi, Aero city----

	venue	freq
0	Hotel	0.35
1	Hotel Bar	0.06
2	Indian Restaurant	0.06
3	Metro Station	0.03
4	Punjabi Restaurant	0.03

Top 10 venues for each locality:

Out[113]:

	Locality	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	ARSS Mall, Paschim Vihar	Indian Restaurant	ATM	Market	Pizza Place	Multicuisine Indian Restaurant	Coffee Shop	Sandwich Place	Chinese Restaurant
1	Adchini	Café	Indian Restaurant	Pub	Hotel	Women's Store	Fried Chicken Joint	French Restaurant	Food Truck
2	Aditya Mega Mall, Karkardooma	Indian Restaurant	Pizza Place	Shopping Mall	Multiplex	Café	Hotel	French Restaurant	Food Truck
3	Aerocity	Hotel	Rental Car Location	Airport Lounge	Coffee Shop	Airport Terminal	Fast Food Restaurant	Falafel Restaurant	Frozen Yogurt Shop
4	Aggarwal City Mall, Pitampura	Indian Restaurant	Pizza Place	Department Store	Food	Fast Food Restaurant	Falafel Restaurant	Frozen Yogurt Shop	Fried Chicken Joint

Clustering:

K-means to cluster the locality 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.

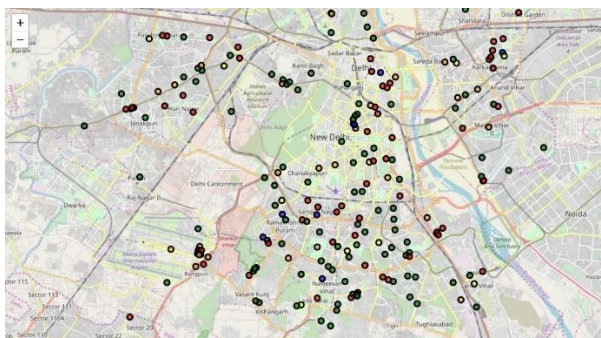
Out[96]:

	Locality	Lat	Lng	No_of_Restaurant	Cuisines	Agg_Rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue
0	ARSS Mall, Paschim Vihar	28.668945	77.101544	1	North Indian, South Indian, Chinese, Mithai, F...	3.100000	Average	117	0	India Restaurant
1	Adchini	28.537063	77.197608	13	Fast Food, North Indian, Seafood, Continental,...	3.292308	Average, Good, Poor, Very Good	1560	2	Café
2	Aditya Mega Mall, Karkardooma	28.656131	77.301266	4	Finger Food, North Indian, Seafood, Pizza, Fast...	3.275000	Average, Good	434	0	India Restaurant
3	Aerocity	28.553077	77.104270	2	Fast Food, Italian, Pizza, North Indian, Cont...	3.200000	Average	59	4	Hot
4	Aggarwal City Mall, Pitampura	28.690020	77.134650	3	North Indian, Chinese, Street Food, Mithai, No...	3.033333	Average	126	0	India Restaurant

RESULT

Cluster maps:

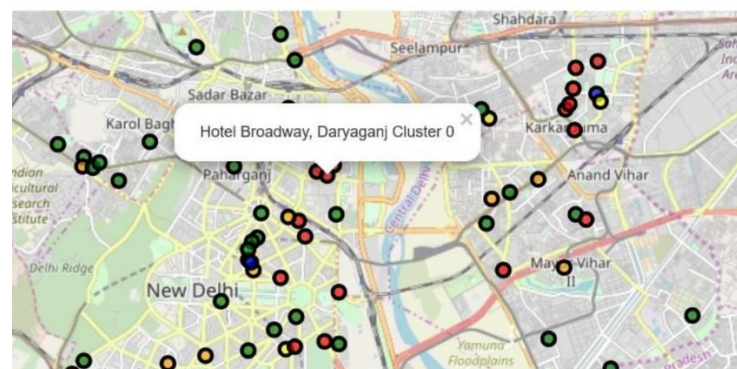
Final cluster map:



There are total 5 clusters.

- **Cluster – 0** - ●
- **Cluster – 1** - ●
- **Cluster – 2** - ●
- **Cluster – 3** - ●
- **Cluster – 4** - ●

CLUSTER – 0 ●



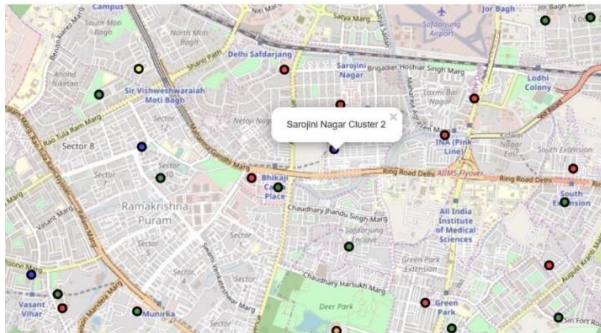
Cluster 0 is defined in red color.

CLUSTER – 1 ●



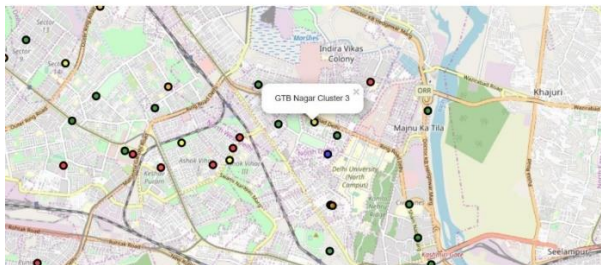
Cluster 1 is defined in green color.

CLUSTER – 2



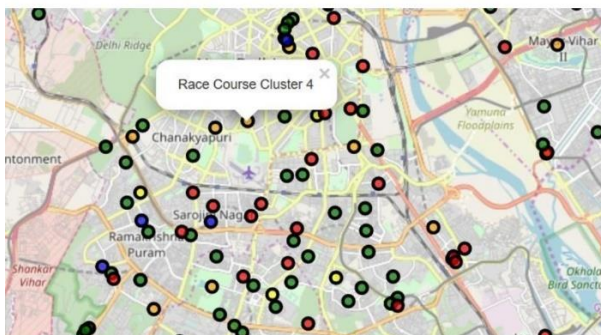
Cluster 2 is defined in blue color.

CLUSTER – 3



Cluster 3 is defined in yellow color.

CLUSTER – 4



Cluster 4 is defined in orange color.

- Chanakyapuri, Pitampura, Safdarjung are some of the best neighbourhoods for Chinese cuisine.
- Panchsheel Park, Nehru place have the best Chinese Restaurant.
- Connaught Place, Rajouri Garden, Malviya Nagar are the best places for edible people.

- Greater Kailash, Firozeshah Rd, Saket have the best restaurants in New Delhi.



Cluster 0: It is most recommended for Indian Restaurants.



Cluster 1: It is most recommended for Hotels and nightclub.



Cluster 2 and Cluster 4: It is most recommended for Fast food.



Cluster 3: It is most recommended for the cafe and pizza.

You can download and use the [New Delhi Clustered Map](#) (HTML File) to view the coolest places in New Delhi.

CONCLUSION:

Clustering is the process to discover similarities between any set of data. Its application areas are data mining, pattern recognition, characterizing web document, grouping genes and proteins that have similar functionality, grouping geographical data like earthquakes and so on. In this paper, various clustering techniques along with their comparative study has been discussed. We can form a hypothesis that the most widely used and studied algorithm is the k-means algorithm. Many algorithms are its variants. A lot of research in this clustering field has emphasized on numerical data sets and only a small number of techniques are available for categorical and other databases.

- Chanakyapuri, Pitampura, Safdarjung are some of the best neighborhoods for Chinese cuisine.
- Panchsheel Park, Nehru place have the best Chinese Restaurant.

- Connaught Place, Rajouri Garden, Malviya Nagar are the best places for edible people.
- Greater Kailash, Firozeshah Rd, Saket have the best restaurants in New Delhi.

REFERENCES

1. Cominetti, Matzavinos, Samarasinghe "DiffFUZZY: 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, Hong Kong.
4. I.K..Ravichandra Rao, "Data Mining and Clustering Techniques", 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. Rousseeuw, 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, Adensity-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of KDD'96, 1996.