

Evaluation of Customer Ratings on Restaurant by Clustering Techniques using R

Ankita Chopra¹, Dr. M. L. Saini²

Assistant Professor¹, Professor²

Department Of Information Technology

Jagan Institute of Management Studies, Rohini, New Delhi¹

Poornima University, Jaipur²

Corresponding author's email id: ankita.chopra@jimsindia.org¹

Abstract

In today's modern times food and lifestyle has become integral part of human system. People today aspire for good day at work and sumptuous and delicious food to eat at the end of the day. Hospitality sector has come up in a big way and in this business serving good food with great ambience has become mandatory to attract customers. So, in this paper we try to collect data from a restaurant in Bangalore and evaluate its popularity based on ratings given by customers. In this paper we use K- Means clustering techniques to cluster the popular restaurants. This analysis would help people to choose restaurants for better food and ambience

Keywords: Food, Clustering, Data Science, Hospitality, K-Means

INTRODUCTION

In today's highly competitive world it is very important for every business sector to retain their customers. Customers are very much aware of their requirements and how best it can get solved. Hospitality business has grown in leaps and bounds over a decade. Customers today would like to explore restaurant depending on the rating

given by customers who visited the restaurant earlier. In this paper we try to analyze customer review data on restaurants of Bangalore in order to find out the most popular restaurant in the city based on ratings given. Data science can be used to develop a model which helps in analyzing the best restaurant from the data.

The algorithm used for analysis is K-means clustering .

Data Mining, also popularly known as knowledge discovery in databases (KDD), refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. While data mining and knowledge discovery (or KDD) are frequently treated as synonyms, data mining is actually part of the knowledge discovery process. The knowledge discovery in databases process comprises of a few steps leading from raw data collections to some form of new knowledge. The iterative process consists of following steps. [12]

- **Data Cleaning:** Also known as data cleaning, it is a phase in which noise data and irrelevant data are removed from the collection.
- **Data Integration:** At this stage, multiple data sources, often heterogeneous, may be combined in a common source.
- **Data Selection:** At this step, the data relevant to the analysis is decided on and retrieved from the data collection.

- **Data Transformation:** Also known as data consolidation, it is a phase in which the selected data is transformed into forms appropriate for the mining procedure.
- **Data Mining:** It is the crucial step in which clever techniques are applied to extract patterns potentially useful.
- **Pattern Evaluation:** In this step, strictly interesting patterns representing knowledge are identified based on given measures.
- **Knowledge Representation:** Is the final phase in which the discovered knowledge is visually represented to the user. This essential step uses visualization techniques to help users understand & interpret the data mining results.

DATA MINING TECHNIQUES

Various algorithms and techniques like classification, clustering, regression, artificial intelligence, neural networks, association rules, decision trees, genetic algorithm, nearest neighbour method etc. are used for knowledge discovery from databases. I Association Association is one of the best known data mining technique. In association, a pattern is discovered

based on a relationship of a particular item on other items in the same transaction. For example, the association technique is used in market basket analysis to identify what products their customers frequently purchase together. Based on this data businesses can have corresponding market campaign to sell more products to make more profit. [12]

CLASSIFICATION

Classification is a data mining technique based on machine learning basically classification is used to classify each item

in a set of data into one predefined set of classes or groups. Classification Techniques

- Regression
- Decision Trees
- Neural Networks

Clustering Clustering is “the process of organizing objects into groups whose members are similar in some way”

- Hierarchical Methods
- Partitioning Methods
- Model based clustering methods

DATA SET & R CODE

Customer Review Data

Name of Restaurant	Rating	Votes
Jalsa	4	775
Spice Elephant	4	787
San Churro Cafe	3	918
AddhuriUdupiBhojana	4	88
Grand Village	4	166
Timepass Dinner	3	286
Rosewood International Hotel - Bar & Restaurant	5	8
Onesta	5	2556

Penthouse Cafe	3	324
Smacznego	3	504
Cafe Shuffle	4	402
The Coffee Shack	4	150
Caf-Eleven	1	164
San Churro Cafe	4	424
Cafe Vivacity	2	918
Catch-up-ino	3	90
Kirthi's Biryani	2	133
T3H Cafe	1	144
360 Atoms Restaurant And Cafe	3	93
The Vintage Cafe	4	13
Woodee Pizza	5	62
Cafe Coffee Day	2	180
My Tea House	3	28
Hide Out Cafe	4	62
CAFE NOVA	4	31
Coffee Tindi	4	11
Sea Green Cafe	1	75
Cuppa	5	4
Srinathji's Cafe	5	23

Redberrys	1	148
Foodiction	5	219
Sweet Truth	4	506
Ovenstory Pizza	1	35
Faasos	1	172
Behrouz Biryani	4	415
Fast And Fresh	4	230
Szechuan Dragon	3	91
Empire Restaurant	5	1647
MaruthiDavangere Benne Dosa	2	4884

APPLICATION OF DATA MINING TECHNIQUE:

Clustering K –Means Model

On the above data set the following R code is implemented:

```
getwd()
setwd("C:/Users/ankita/Documents")
data=read.csv("dataset.csv")
print(data)
cor.test(data$Rating,data$Votes)
ds=data$Votes
sort(ds)
ds1=data$Rating
sort(ds1)
km=kmeans(ds,5,40)
km
```

```
plot(ds,col=km$cluster)
```

```
hist(data$Rating)
```

```
plot(data$Rating,type="o",col="blue",xlab
= data$Votes)
```

MEASURES OF PERFORMANCE

In this section we try to analyze the code written above.

Firstly as the data set contains 3 attributes they are name of restaurant, rating and votes given by customer. We first build 5 clusters depending on votes and rating given by customers and the result is illustrated in plot as follows:

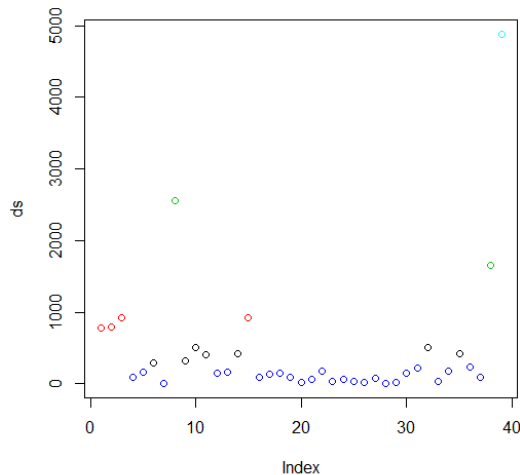


Figure 1 Showing Clusters as per ratings

The above plot clearly shows that more number of points in cluster is for rating given as 4.

Secondly correlation test is conducted between rating and votes given by customers and the result are as follows:

data: data\$Rating and data\$Votes

$t = 0.005578$, $df = 37$, $p\text{-value} = 0.9956$

Alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.3146911 0.3163425

Sample estimates:

cor

0.0009170227

Correlation results also prove that there is strong correlation between rating parameter and votes parameter given by customers.

Thirdly as it is deduced that rating and votes are strongly correlated plotting is done further to understand and analyse most popular restaurant depending on rating.

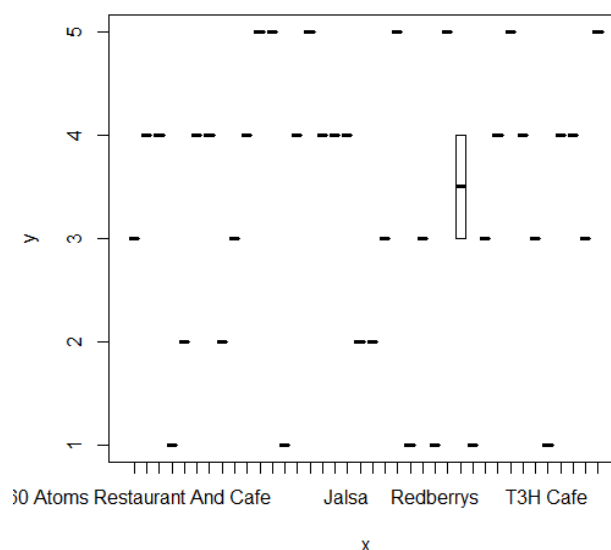


Figure 2 showing most popular restaurant as per rating

Lastly we plot a line chart

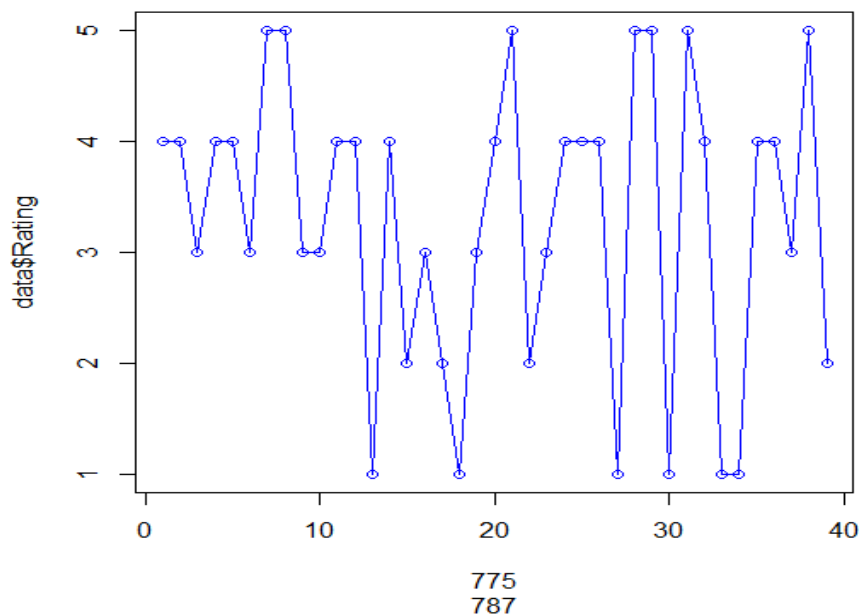


Figure 3 showing most popular restaurant as per votes

CONCLUSIONS AND FUTURE SCOPE

In this paper we developed a clustering model which deduced the following things

1. Rating of restaurant and votes given by customers after visiting the restaurant are strongly correlated with each other.
2. The restaurant with highest number of common rating is very popular among customers so, among the database T3H café and jalsa restaurant are very popular.

The future scope of the model lies in the idea of studying it with few more

parameters of customer reviews and analyzing it with classification algorithm and then comparing the accuracy of both to understand which model works best to understand customer pattern in hospitality business.

REFERENCES

1. <http://searchbusinessanalytics.techtarget.com/essentialguide/tapping-the-potential-of-social-media-analytics-tools>
2. <http://onlinelibrary.wiley.com/doi/10.1111/jcc4.12029/full>
3. <http://blog.sagecrm.com/detail.php?id=27782/14-experts-on-the-biggest-customer-service->

- challenges-faced-by-today-39-sbusinesses
4. P.mishra, A. bhattacharjee, capturing the voice of the employee: enterprise social media monitoring and analytics, white paper,tcs.
 5. Social media for utilities: developing a satisfying customer experience, cognizant 20-20 insights, 2013.
 6. H.chen, r. h. l. chiang,v.c. storey, business intelligence and analytics: from big data to big impact, mis quarterly vol. 36 no. 4/december 2012
 7. Tuning in to the emotions of the capital markets with sentiment analysis, tcs. r.shullich, risk assessment of social media, the sans institute,2012
 8. bandra, f.ioras, k.maher, “cyber security concerns in elearning education,” international conference of education, research and innovation seville, spain. proceedings of iceri2014 conference, nov 17-19, 2014.
 9. N. barik, dr. s. karforma, “risks and remedies in e-learning system,” international journal of network security & its applications (ijnsa), vol. 4, no. 1, pp. 51-59, jan 2012.
 10. Cristo´balromero *, sebastia ´nventura, enriquegarcia ”data mining in course management systems: moodle case study and tutorial, department of computer sciences and numerical analisis, university of co´rdoba,14071co´rdoba,spainnov17-19, 2014
 11. Ankita chopra,” quantitative analysis of dairy product packaging with the application of data mining techniques, ”international journal of computer science and information technology,issn:0975-9646,vol 7,no 2,mar-apr, 2016.2009,
 12. “Power to the people – social media tracker wave,” <http://universalmccann.bitecp.com/wave4/wave4.pdf>, universal mccann.