Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 10, October-2018

Segmentation of Mobile Customers using Data Mining Techniques

Md. Mahfuz Reza
Assistant Professor
Department of Computer Science and Engineering
Mawlana Bhashani Science and Technology University
Santosh, Tangail-1902, Bangladesh

Sajedun Nahar
Student
Department of Computer Science and Engineering
Mawlana Bhashani Science and Technology University
Santosh, Tangail-1902, Bangladesh

Tanya Akter
Student
Department of Computer Science and Engineering
Mawlana Bhashani Science and Technology University
Santosh, Tangail-1902, Bangladesh

***Abstract* - In today's competitive world, to keep customers satisfied is a key to success for telecommunication companies. Data mining techniques are more preferable for discovering the customer's attributes as well as their needs which is possible by segmenting their behaviours. Segmentation is the process of developing meaningful customer groups that are similar based on individual account characteristics and behaviours. Using K-means clustering, the paper proposes a resolution of customer segmentation for the telecommunication company. The prime objectives were to group customers using their behavioural characteristics and provide services according to the group.**

Keywords: Data base of mobile customers, Data mining, K-means clustering, segmentations and services

I. INTRODUCTION

Telecom industry is a typical data-intensive industry, in which data mining technologies can be used to obtain useful knowledge to provide customers with better services and find more commercial opportunities[2].Customer satisfaction and attraction are one of the most significant goals in top level leading companies today. It will directly impact on company's revenue and income .Customers profitability is the profit that the company makes from serving a customer or customer group over a specified period of time [3]. Customer segmentation is a term used to describe the process of dividing customers into homogeneous groups on the basis of common attributes [16]. The customers within the same group have greatest similarity. Telecommunication companies utilize data mining to improve their marketing efforts, identify fraud, and provide service to the customer [6].

Data mining refers to extracting or mining knowledge from large amount of data. Many other terms carry a similar or slightly different meaning to data mining, such as knowledge mining fromdata, knowledge extraction, data or pattern analysis, data archaeology and data dredging.

Commonly used data mining techniques includes association analysis, classification and prediction, cluster analysis, outlier analysis and evolution analysis. Among them, the cluster analysis can be used to solve the problem of customer grouping [7].

Clustering basically deals with grouping of objects such that each group consists of similar or related objects. The main idea behind clustering is to maximize the intra-cluster similarities and minimize the inter cluster similarities. In this thesis paper, we use K-means clustering technique to segment the customers [2] [5]. This is one of the most common and effective method to classify data because of its simplicity and ability to handle voluminous data sets. Generally, it accepts the number of clusters and the initial set of centroids as parameters. The distance of each item in the data set is calculated with each of the centroids of the respective cluster. For grouping the items of a data set using K-Means clustering is calculating the distance of the point from the chosen mean. This distance is usually the Euclidean Distance.

II. LITERATURE REVIEW

In telecom sector, customer clustering or segmentation is one of the most significant methods used in studies of marketing. To arrive a better justification regarding of customer analysis for providing offers, it is relevant to make a survey of related literature.

Some research about segmentation for customers has been developed. Ours is based on the discussion, proposed by S.M.H. Jansen, useddifferent clustering techniques to segment the customers and support vector machine to profile the segmented customer (2007).

KonstantinosTsiptsis and AntoniosChorianopoulos discussed the customer segmentation in telecommunication on the basis of user behaviour using two approaches of segmentations(2009).One is behavioural analysis and another is value-based segmentation. They used all the available usage data to reveal the natural groupings in their customer base. The behavioural segmentation implementation included the application of a data reduction technique (PCA) to reveal the distinct dimensions of information, followed by a clustering

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 10, October-2018

technique to identify the segments. And value-based segmentation relies only on a single field. It does not need the application ofa data mining algorithm either. It only involves a simple sorting of records according to aprofitability index and an assignment to corresponding groups.

In 2017 3rd International Conference on Science and Technology - Computer (ICST) , AndryAlamsyah and BellaniaNurriz performed a business analysis "Monte Carlo Simulation and Clustering for Customer Segmentation in Business Organization" to support customer segmentation case . They applied clustering methods in one of the branches of Indonesia Telecommunication Company. They wanted to show a simulation to generate customer's income data for "Telkom Indonesia Makassar city".Furthermore, the simulation result data is used to supportcustomer segmentation analysis using the K-Means clustering.

III. THEORITICAL BACKGROUND

A.Data Mining

Data mining is the process of searching and analyzing data in order to find implicit, but potentially useful information. It is a powerful tool, helpful to companies as it predicts customers[1]. There are some basic data mining tasks such as association rules, sequential pattern, clustering and classification.

B. Clustering

The objective of cluster analysis is the organization of objects into groups, according to similarities among them[10]. The main idea behind clustering is to maximize the intra-cluster similarities and minimize the inter cluster similarities. K-Means algorithm is one of the common clustering processes based on centroid model.K-Means algorithmis a classical algorithm to solve the clustering problem [16]. The measure for the case in the cluster is represented by the meanvalue.

K-means works as follows:

1. Select the number of cluster. Let this number be K.
2. Pick K seeds as centroid of thecluster. The seeds may be picked randomly.
3. Compute the Euclidean distance of each object in the set from each of the centroids.
4. Allocate each object to the cluster it is nearest to based on distances computed in the previous step.
5. Compute the centroid of clusters by computing the means of the attribute values of cluster.
6. Check if the stopping criteria has been met, if yes go to stop.
7. else go to 3.

C. Segmentation

Segmentation is a process to divide customersof a consumer or business market in groups based on someshared characteristics. There are various aspects on whichcustomer segmentation can be done such as demographic, behavioral, geographic and so on[3].

IV. RESEARCH METHODOLOGY

This section describes in detail the research methodology including data collection, preparation, cluster analysis, segmentation, profiling, customer identification and service providing.

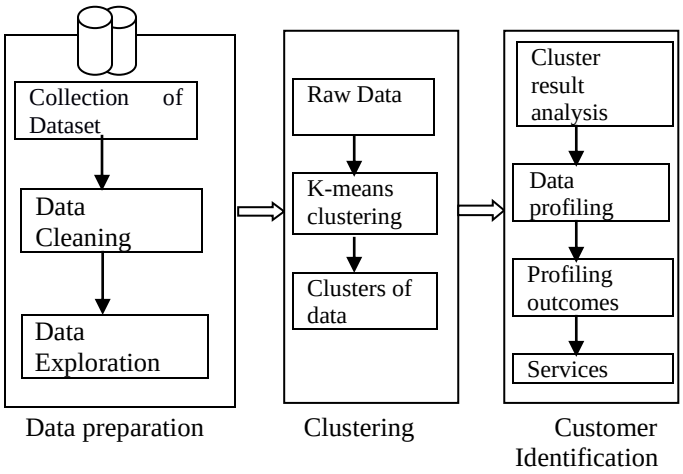


Fig. 1: Flow chart of customer segmentation

The first process is collecting the data from a telecom industry.

The attributes in our data set which is given below [6]:

TABLE I. DATASET ATTRIBUTES

S.No.	Attribute names
01	Phone_Number
02	Age
03	Sex
04	Network Type
05	Daytime_Mins
06	Evening_Mins
07	Night_Mins
08	Incoming_Mins
09	Outgoing_Mins
10	In-SMS
11	Out_SMS
12	Calls_Roaming
13	Internet
14	Hour

After we get the data,it needs to be cleaned and prepared the data.The pre-processing step contains several activities such as noise reduction,data cleaning, data integration and data reduction to get a better data form. The third process is segmenting the customers using cluster analysis.K-means is one of the most important and commonly used method fordividing the dataset into several clusters that requested[5]. The fifth process is analysing the result of the customers. Finally, we provide services to these customers.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 10, October-2018

V. EXPERIMENT & RESULT

We analyse fifteen days historical data of anonymized coded dataset, for protecting customer and company privacy. Implementation is done by using rstudio and oracle 11gR2 software.

A. Checking missing value:

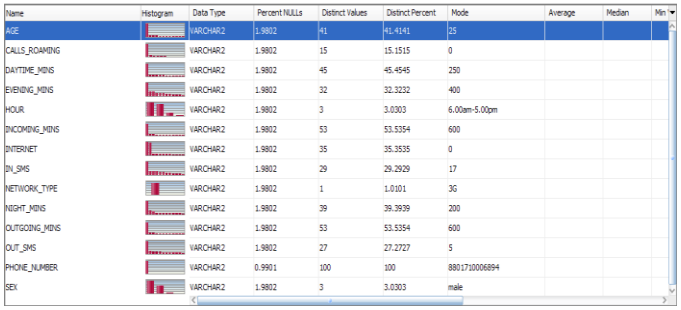
We will check whether any missing value exists or not in our data set. We have used R language for calculating missing value.

Phone_Number	Age	sex	Network_Type	Daytime_Mins
0	0	0	0	0
Evening_Mins	**Night_Mins**	**Incoming_Mins**	**outgoing_Mins**	**In_SMS**
0	0	0	0	0
Out_SMS	**Calls_Roaming**	**Internet**	**Hour**	
0	0	0	0	

Fig. 2: Calculation of missing value

B.Data Exploration:

Before clustering, data exploration should be done. We use oracle data miner (ODM) to explore the data.



Name	Histogram	Data Type	Percent NULLs	Distinct Values	Distinct Percent	Mode	Average	Median	Min
AGE		VARCHAR2	1.9802	41	41.4141	25			
CALLS_ROAMING		VARCHAR2	1.9802	15	15.1515	0			
DAYTIME_MINS		VARCHAR2	1.9802	45	45.4545	250			
EVENING_MINS		VARCHAR2	1.9802	32	32.3232	400			
HOUR		VARCHAR2	1.9802	3	3.0303	6.00am-5.00pm			
INCOMING_MINS		VARCHAR2	1.9802	53	53.5354	600			
INTERNET		VARCHAR2	1.9802	35	35.3535	0			
IN_SMS		VARCHAR2	1.9802	29	29.2929	17			
NETWORK_TYPE		VARCHAR2	1.9802	1	1.0101	3G			
NIGHT_MINS		VARCHAR2	1.9802	39	39.3939	200			
OUTGOING_MINS		VARCHAR2	1.9802	53	53.5354	600			
OUT_SMS		VARCHAR2	1.9802	27	27.2727	5			
PHONE_NUMBER		VARCHAR2	0.9901	100	100	8801710006894			
SEX		VARCHAR2	1.9802	3	3.0303	male			

Fig. 3: Data exploration

C. Structure of Dataset:

Information of all attributes are included in this section.

```
'data.frame':   99 obs. of  15 variables:
 $ Phone.Number : num  8.8e+12 8.8e+12 8.8e+12 8.8e+12 8.8e+12 ...
 $ Age          : int  20 45 33 67 56 53 34 26 45 26 ...
 $ sex          : Factor w/ 3 levels "female","femlae",..: 1 3 3 3 1 3 3 3 3 1 ...
 $ Network.Type : Factor w/ 1 level "3G": 1 1 1 1 1 1 1 1 1 1 ...
 $ Daytime.Mins : int  250 420 580 210 280 300 580 1200 299 1160 ...
 $ Evening.Mins : int  220 500 450 250 400 430 450 700 400 420 ...
 $ Night.Mins   : int  300 150 200 59 90 70 650 1000 90 600 ...
 $ Incoming.Mins: int  520 470 380 400 520 400 680 1600 389 1160 ...
 $ outgoing.Mins: int  250 600 850 119 250 400 1000 1300 400 1020 ...
 $ In.SMS       : int  54 24 21 11 17 15 21 23 30 50 ...
 $ Out.SMS      : int  32 0 4 0 2 5 4 12 33 30 ...
 $ Calls.Roaming: int  0 0 30 59 0 0 30 80 0 50 ...
 $ Internet     : int  1024 0 0 0 0 400 0 5000 456 9067 ...
 $ Hour         : Factor w/ 3 levels "5.00pm-8.00pm",..: 3 1 2 1 1 1 2 2 1 2 ...
```

Fig. 4: Structure of Dataset

D. Cluster Analysis:

In this research, K-means clustering is used to group the customer for further segmentation.

K-means is a centroid based clustering in which the notion of similarity is derived by how close a data point is to the centroid of the cluster. The item is then assigned to the cluster with which the distance of the item.

```
K-means clustering with 3 clusters of sizes 17, 60, 22

Cluster means:
  Phone.Number      Age Daytime.Mins Evening.Mins Night.Mins Incoming.Mins
1 8.801750e+12 34.70588     537.1765     423.9412   298.2353       617.8235
2 8.801722e+12 37.35000     563.2333     445.5667   246.2667       607.1833
3 8.801778e+12 38.50000     541.4545     450.5909   242.5000       580.6818
  outgoing.Mins   In.SMS  Out.SMS Calls.Roaming Internet
1      635.6471 33.17647 28.70588      28.52941 1974.412
2      639.7167 28.31667 18.10000      27.25000 1589.600
3      644.7727 27.18182 18.90909      31.50000 1475.045
```

Fig. 5: Centroid information

We segment the "Hour" attribute into three time periods. This are-

- Attribute Daytime. Mins denotes the period 6.00am-5.00pm.
- Attribute Evening.Mins denotes the period 5.00pm-8.00pm.
- Attribute Night. Mins denotes the period 8.00pm – 6.00am.

E. Profiling

We profile the customers using hour attribute. This hour attribute will provide segments. The figure given below distributes the hour profile of each cluster.

	1	2	3
5.00pm-8.00pm	8	25	8
6.00am-5.00pm	7	30	11
8.00pm-6.00am	2	5	3

Fig. 6: The hour profile of each cluster

F. *Relativity*

Relationship among Daytime.Mins, Evening.Mins and Night. Mins customers are given below:

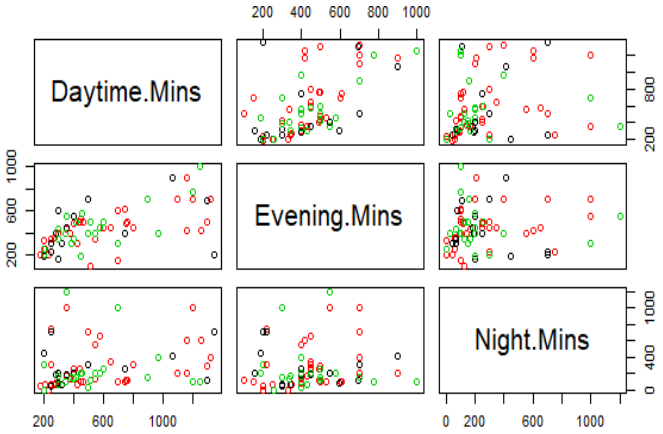


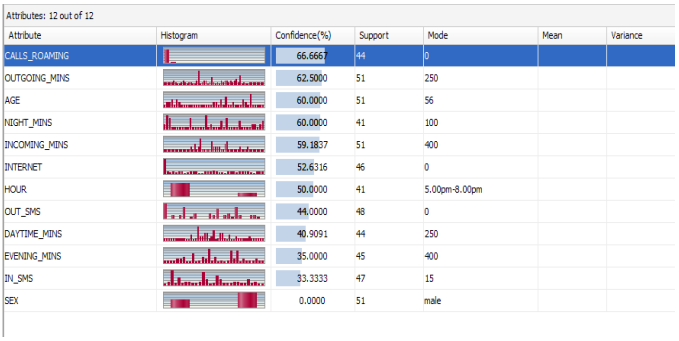
Fig. 7: Relativity in Daytime.Mins, Evening.Mins, Night.Mins

By examining the values of the each cluster we can determine the profitability of the customer. By analysing above figure 6 we can determine cluster 2 is the most profitable customer. Cluster 1 includes low profitable customers. So, we can decide:

Cluster 2: High profitable customers.
Cluster 3: Profitable customers.
Cluster 1: Low profitable customers.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 10, October-2018

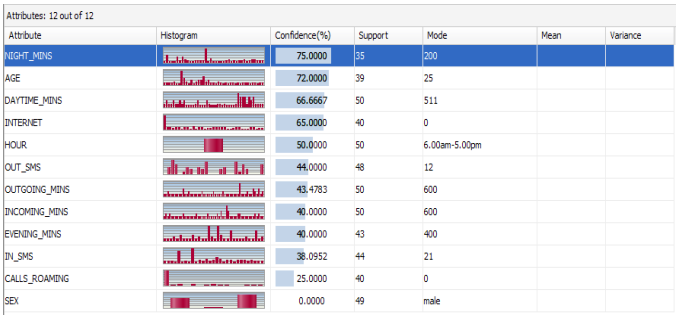
Cluster 2 contains both Daytime. Mins and Evening. Mins attributes.



Attributes: 12 out of 12

Attribute	Histogram	Confidence(%)	Support	Mode	Mean	Variance
CALLS_ROAMING		66.6667	44	0		
OUTGOING_MINS		62.5000	51	250		
AGE		60.0000	51	56		
NIGHT_MINS		60.0000	41	100		
INCOMING_MINS		59.1837	51	400		
INTERNET		52.6316	46	0		
HOUR		50.0000	41	5.00pm-8.00pm		
OUT_SMS		44.0000	48	0		
DAYTIME_MINS		40.9091	44	250		
EVENING_MINS		35.0000	45	400		
IN_SMS		33.3333	47	15		
SEX		0.0000	51	male		

Fig. 8: Information of Evening. Mins user



Attributes: 12 out of 12

Attribute	Histogram	Confidence(%)	Support	Mode	Mean	Variance
NIGHT_MINS		75.0000	35	200		
AGE		72.0000	39	25		
DAYTIME_MINS		66.6667	50	511		
INTERNET		65.0000	40	0		
HOUR		50.0000	50	6.00am-5.00pm		
OUT_SMS		44.0000	48	12		
OUTGOING_MINS		43.4783	50	600		
INCOMING_MINS		40.0000	50	600		
EVENING_MINS		40.0000	43	400		
IN_SMS		38.0952	44	21		
CALLS_ROAMING		25.0000	40	0		
SEX		0.0000	49	male		

Fig. 9: Information of Daytime.mins user

By understanding profitability and attributes of customers, companies can make decisions to improve their services.Companies need to serve better service by handling above discussed customer category. After grouping the customers then we provide services among them.

Services provided for Daytime users:

- On-net call time offers.
- Bundle offers.
- Bonus offers.
- Offers for news, games facility.

Services provided for Evening time users:

- FnF offers.
- Special call rate offers.
- SMS bundle offers.
- Provides with new technology services like internet offers.

We can't ignore medium and low profitable customers. Because they are part of the companies' profit. One day they can be loyal customer. So, companies have to concern about them.

VI. CONCLUSION & FUTURE WORK

A. Conclusion

Latterly, the mobile telecommunication marketplace is highly competitive. Increasing the number of customers is the main challenge in modern telecommunication industry[4]. In this paper, we have shown that through the use of customer segmentation,a telecommunication company can easily attract its customers with right products and services[6]. This also helps in offering packages, offers and bundles for customers. So, companies must realize the grandeur of customer segmentation and profiling the customer's behaviour to achieve better results by narrowing customer segments.The cluster analysis is able to solve customer segmentation problem[18]. This paper adopts the K-means clustering method to resolve a analysis of telecom customer segmentation. Practical results indicate that the analysis of customer segmentation for telecom sector is effective and successful[2]. The business objective was to group customers in terms of their behavioural characteristics and provides services according to the group considering which customers are profitable for the company.In other words, we theoretically discuss about the utilization of data mining algorithm for serving with suitable offers to the customers.

B.Future work

For future research we can predict the risk customer using association rules..In future, we can find accuracy of all customers using churn prediction. Also we can include studying the performance of clustering with applying the behaviour of revenue.

REFERENCES

[1] S.M.H. Jansen, "*Customer Segmentation and Customer Profiling for a Mobile Telecommunications Company Based on Usage Behavior: A Vodafone Case Study*,"pp 8-13, 17July 2007.

[2] CaiQiuru, Lua Ye, Xi Haixu, Liu Yijun, Zhu Guangping. "*Telecom Customer Segmentation Based on Cluster Analysis*,"inprceedings of International Conference on Computer Science and Information Processing (CSIP), 2012:IEEE.

[3] A. S. M. Shahadat Hossain , "*Customer Segmentation using Centroid Based and Density Based Clustering Algorithms*,"in proceedings of the 3rd International Conference on Electrical Information and Communication Technology (EICT), 7-9 December 2017, Khulna, Bangladesh:IEEE.

[4] AnahitaNamvar, Mehdi Ghazanfari, Mohsen Naderpour, "*A Customer Segmentation Framework for Targeted Marketing in Telecommunication*,"in proceedings of 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE),2017:IEEE.

[5] AndryAlamsyah, BellaniaNurriz , "*Monte Carlo Simulation and Clustering for Customer Segmentation in Business Organization*,"in proceedings of the 3rd International Conference on Science and Technology- Computer (ICST),2017:IEEE.

[6] Salar Masood, Moaz Ali, Faryal Arshad, Ali Mustafa Qamar, Aatif Kamal, Ahsan Rehman, "*Customer Segmentation and Analysis of a Mobile Telecommunication Company of Pakistan using Two Phase Clustering algorithms*," 2013: IEEE.

[7] ZHANGYihua, "*Vip Customer Segmentation Based on Data Mining in Mobile-communications Industry*," in proceedings of The 5th International Conference on Computer Science & EducationHefei, China. August 24–27, 2010: IEEE.

[8] HasanZiafat, Majid Shakeri, "*Using Data Mining Techniques in Customer Segmentation*," in proceedings of HasanZiafat Int. Journal of Engineering Research and Applications, Vol. 4, Issue 9(Version 3), September 2014, pp.70-79.

[9] Prof. TejalUpadhyay, "*Customer Profiling and Segmentation using Data Mining Techniques*,"vol.7, no.2, march 2016-sept 2016,pp 65-67 [online] Available:www.csjournalss.com.

[10] Mali, K., "*Clustering and its validation in a symbolic framework. Patt. Recogn*," Lett., vol. 24 (2003), pp. 2367-2376.

[11] Anshul Arora ,Dr. Rajan Vohra, "*Segmentation of Mobile Customers for Improving Profitability Using Data Mining Techniques*," in the proceedings of (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , pp 5241-5244,2014.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 10, October-2018

[12] HasithaIndika Arumawadu1, R. M. KapilaTharanga Rathnayaka2,3, S. K. Illangarathne4, "*Mining Profitability of Telecommunication Customers Using K-Means Clustering*," in Journal of Data Analysis and Information Processing, 2015, vol.3, pp 63-71,[online] Available:http://www.scirp.org/journal/jdaip,http://dx.doi.org/10.4236/jdaip.2015.33008 , [Accessed in August 2015 in SciRes].

[13] Emilia Mattila, "*Behavioral Segmentation of Telecommunication Customers*" inMaster of Science ThesisStockholm, Sweden 2008.

[14] SabriSerkanGüllüoğlu, "*SEGMENTING CUSTOMERS WITH DATA MINING TECHNIQUES*,", 2015:IEEE.

[15] KonstantinosTsiptsis ,AntoniosChorianopoulos, "*Segmentation Applications in Telecommunications*" in Data Mining Techniques in CRM: Inside Customer Segmentation,pp 291-332,2009.

[16] Ye, Xi Haixu, Liu Yijun,CaiQiuru,Yu Zhi-min. "*Telecom Customer Segmentation Based on Cluster Analysis*,"inprceedings of International Conference on Computer Science and Information Processing (CSIP),2012:IEEE.

[17] Jinghua Zhao, Wenbo Zhang, Yanwei Liu, "*Improved K-Means Cluster Algorithm in Telecommunications Enterprises Customer Segmentation*" ,2010:IEEE.

[18] ZhangXiaozhou, Zhang Yonghong, "Telecom CustomerMarketing [M]" Beijing: Posts & Telecom Press, 2008.