

## **Mining The Usage of Data Plans For Mobile Customers Using Clustering Algorithm**

**P.B. Pankajavalli<sup>1</sup>, Manimozhi B<sup>2</sup>, Narendhiran B<sup>3</sup>**

<sup>1</sup> Assistant Professor, <sup>2,3</sup> PG Student

*Department of Computer Applications, PSG College of Technology, Coimbatore -  
641004, Tamilnadu, India.*

*E-mail: [pankajavalli.g@gmail.com](mailto:pankajavalli.g@gmail.com), [pbp@mca.psgtech.ac.in](mailto:pbp@mca.psgtech.ac.in)*

### **Abstract**

Introducing new strategies in marketing makes a sound customer relationship. Since marketing management is towards the customers, achieving success depends on the loyalty and valued customer relationship. Therefore it is essential to evolve new strategies based on the customer preferences.

The objective of this paper is to propose a model for the clustering the customer based on the parameters. This clustering enables mobile service providers to enhance their business which will help the service providers to go for contract with the mobile manufacturers.

### **Introduction**

Customer Relationship Management (CRM) is considered as major factor in competitive business environment. Some potential benefits of CRM areas follows:

- (1) Increased customer retention and loyalty,
- (2) Higher customer profitability,
- (3) Creation value for the customer,
- (4) Customization of products and services,
- (5) Lower process, higher quality products and services.

Telecom industry is one of the fastest growing industries. So, CRM plays vital role in industrial growth. The competition to acquire and retain customers among mobile service providers is essential.

In this paper, clustering of mobile customers is done for analyzing market trend. Clustering is done based on usage patterns. The factors considered for usage pattern includes salary, mobile preferences, usage of data, GPS and social networking, messages per day, gender and age.

### **Significance of The Problem**

- Revenue maximization
- Retaining customers

## Related Work

This study organizes existing customer segmentation methods suggested by few authors.

In [1] the author has applied demographic clustering algorithm to categorize the customers and applied two levels to come out with data categorization.

In [2] the authors has categorized based on some rules, patterns and based on network providers. The networks providers are categorized with names a, b ,c and d.

The authors have categorized the customers based on Life Time Value (LTV). The LTV calculations are done based on the factors like current vale, potential value and the loyalty [3] [4].

## Data Mining and Clustering Methods

Data mining - also known as knowledge-discovery in databases (KDD) is process of extracting potentially useful information from raw data. A software engine can scan large amounts of data and automatically report interesting patterns without requiring human intervention.

Clustering is a typical unsupervised learning technique for grouping similar data points. To assign large number of data points to various small groups with similar and dissimilar properties will be done by clustering algorithms [5]. Sankar Rajagopal (2011) has considered parameters like recency, total customer profit, total customer revenue and top revenue department to categorize the customers.

**Customer Clusters:** Customers clustering is the most important data mining methodologies used in marketing and CRM. Data items are grouped according to logical relationships or consumer preferences.

## Clustering Methods

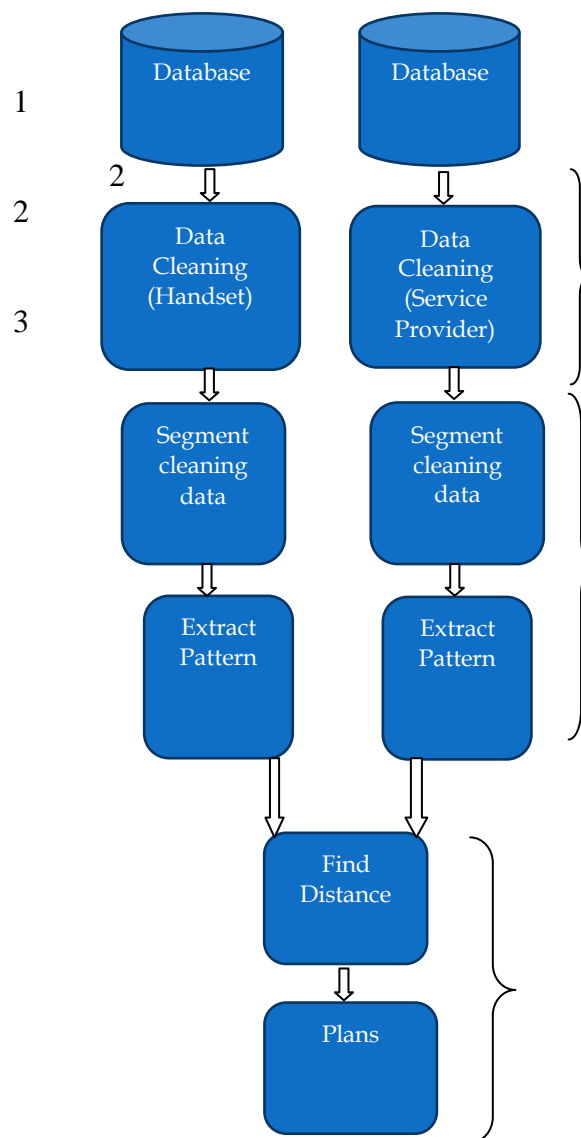
Clustering is unsupervised learning technique for grouping similar data points. A clustering algorithm assigns a large number of data points to a smaller number of groups such that data points in the same group share the same properties while, in different groups, they are dissimilar[6]. Clustering has many applications, including part family formation for group technology, image segmentation, information retrieval, web pages grouping, market segmentation, and scientific and engineering analysis.

In this paper K-means algorithm is applied. When number of variables is larger K-means is computationally faster and it produces tighter clusters. These features motivated to choose k-means clustering approach. It is a non-hierarchical approach for forming good clusters. Here K-means algorithm was used to divide customers into same groups of customers based on customer behavior [7].

## Proposed Architecture

The objective of the proposed system is to develop a model to cluster the customers based on their usage patterns, the activities they indulge in and preferred handsets. The datasets considered for the given model is shown in table 1, 2 and 3. Table 1 and

2 shows few sample handset data considered and table 3 shows the few sample dataset of SMS plans offered to the customers.



**Figure 1:** Framework for clustering the

**Customers**

In Fig1. The classification of the diagram represents:

1. Data Preparation
2. Clustering Process
3. Plans Suggestion

### Data Preparation

This module does the job of extracting all the information from the database. Few sample data are show in the below tables. The data cleaning is done on the gathered data and comes out with the required data.

Few hand set and SMS data plan and sample data are show below:

**Table 1:** Handset data set 1

Apple	Black-berry	Micro-max	Sam-sung	Nokia	Monthly exp.	Avg. min per day	SMS per day	Social Networ king
1	0	0	0	0	4000	50	10	1
0	1	0	0	0	3000	60	15	1
0	0	1	0	0	1000	95	5	1
0	0	0	0	1	1700	95	15	0
1	0	0	0	0	2800	95	30	1
0	0	0	1	0	1000	75	10	1
0	0	1	0	1	700	90	5	0
0	1	0	0	0	1000	60	5	1
0	0	0	1	0	2500	90	5	0

**Table 2:** Handset data set 2

e-mail checking	Avg use of GPS	Use of Online Games	Music and Video download	Current data plan	Age	Yearly income	Gend er
1	1	1	1	7000	20	1500000	1
1	1	1	1	5000	27	1000000	1
1	0	0	0	2000	45	500000	0
1	0	0	1	1000	50	400000	0
1	0	1	0	1000	25	1200000	1
0	0	1	1	1500	25	800000	1
1	0	0	0	500	29	400000	0
1	1	0	0	1000	35	400000	0
1	0	0	1	3000	43	300000	0
1	1	0	0	1000	55	350000	0
1	0	0	1	3000	30	1000000	1
1	1	0	1	8000	29	1500000	1

**Table 3:** SMS data plans

Service Provider	MRP	Validity	Talk Time	Description	SMS per day
<b>Aircel</b>	6	2	0	100 Local/National 100	100
<b>Aircel</b>	19	28	0	750 Free Local SMS. First SMS of day charged at 50p	750
<b>Airtel</b>	49	28	0	100 Local or STD SMS/day	2800
<b>Airtel</b>	92	60	0	100 Local or National SMS/day	6000
<b>BSNL</b>	12	7	0	130 Local/ National SMS free	130
<b>BSNL</b>	32	30	0	385 Local/ National SMS free	385
<b>Idea</b>	18	14	0	50 Local SMS/day	700
<b>Idea</b>	21	30	0	50 Local SMS/day. First SMS of the day charged @ 50p	1500
<b>ReliaCDMA</b>	15	30	0	Local Onnet @ 5p and other @ 15p	3000
<b>ReliaCDMA</b>	19	15	0	2000 Free Local + National SMS - (Max 100 SMS/day)	200
<b>ReliaGSM</b>	15	30	0	Local Onnet @ 5p and other @ 15p	3000
<b>ReliaGSM</b>	33	30	0	400 SMS (Local + National)	400
<b>Tata DoCoMo</b>	16	14	0	100 Local SMS daily (First SMS @ 50p/day)	1400
<b>Tata DoCoMo</b>	26	28	0	100 Local SMS daily (First SMS @ 50p/day)	2800
<b>Tata DoCoMo</b>	38	28	0	100 Local/National SMS daily (First SMS @ 50p/day)	2800
<b>Uninor</b>	7	15	0	25 Local/National SMS Free	25
<b>Uninor</b>	13	28	0	50 Local/National SMS Free	50
<b>Vodafone</b>	27	28	0	100 Local SMS/Day. First SMS of the day charged at 50p/SMS	2800
<b>Vodafone</b>	38	28	0	100 Local and National SMS/Day. First SMS of the day charged at 50p/SMS	2800

### Clustering Process

Based on the selected handset, contract with service plans will be done which will cater to their areas of clusters of people with similar data-usage attributes and mobile preferences. The clusters identified by the system can be targeted by service providers with customized data plans and contracts. For a service provider, having a well-defined idea about customer preferences can help them to bind relationship with existing high value customers and attract new customers with tailored tariffs and promotions.

Customer preferences (datasets) and usage are given as input. It is processed using K-Mean clustering algorithm and number of clusters formed. Each cluster represents different customer segments [8] [9].

The clusters are labeled as High Income Business Users, Office Users, Handset Agnostic Users, Value Maximizers, and New Generation Users.

### **Plans Suggestion**

Distance is calculated from the extracted data. The difference value is ranked as very low, low, average, high and very high. These clusters are compared with segmented customers and ranking for all service providers under each cluster will be shown. Suggestion for all service providers are provided based on ranking under each customer segment.

Depending on the ranking level the suggestion will be given. If the rank level is low, the service provided by the concerned handset is good and when the ranking level is very high, few changes in their plan is required.

These suggestions are used by service providers to have contract with the mobile manufacturers.

### **Results and Discussion**

“Segmentation of customers for contract and data plans” does clustering of customers based on the attributes. The cluster obtained by the clustering technique represents the customers having different usage and preferences. The data plans provided by various service providers help the customers to select the plans of their choice. This enables the service providers to collaborate with mobile handset manufacturers to sell contract based mobile handsets at a discounted price along with new and innovative product offerings to the customers. The suggestions for the offers presented using the system can be used to frame new plans.

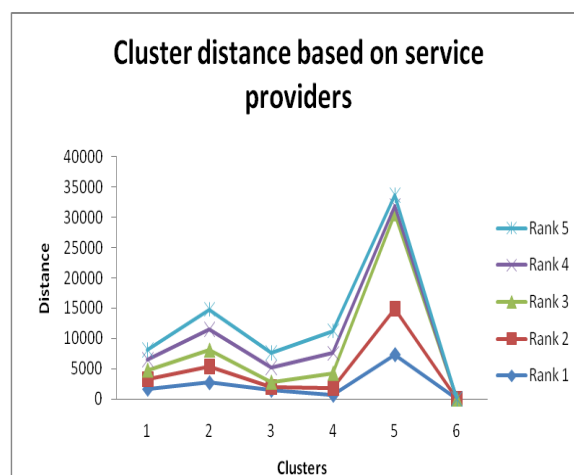
#### **Based on Customer Usage Patterns**

- Cluster 1 ->High Income Business Users
- Cluster 2 ->Office Users
- Cluster 3 ->Value Maximizers
- Cluster 4 ->Handset Agnostic Users
- Cluster 5 ->Online Generation Users.

#### **Based on Data and SMS Plans, Distance Is Ranked As:**

- Rank 1 -> Very low
- Rank 2 ->Low
- Rank 3 ->Average
- Rank 4 ->High
- Rank 5 ->Very high

The distance is calculated based on the handset data and data plan. The distance is compared with the cluster information.



**Figure 2:** Results based on distance

## Conclusion and Future Scope

In today's business environment customer satisfaction is the main mantra. To improve the satisfaction of customers, the customer are segmented based on their usage and their preferences. Various clusters are formed and distance is calculated from the gathered information. Based on the distance the service providers ranking is done. This will enable the service providers to focus on the lacking area and for further introduction of new schemes.

The further work can be extended on segmentation (clustering) using more detailed behavioral data and opportunity identification using association algorithms within the segments discovered.

## References

- [1]. Guangli Niew, Yibing Chen, Lingling Zhanga, and Yuhong Guo, 2012, "Credit card customer analysis based on panel data clustering", International Conference on Computational Science ICCS, Elsevier, pp. 2489 - 2497.
- [2]. Mohammad Ali Farajian & Shahriar Mohammadi, 2010, "Mining the Banking Customer Behavior Using Clustering and Association Rules Methods", International Journal of Industrial Engineering & Production Research, Volume 21, Number 4, pp. 239-245.
- [3]. Su-Yeon Kim , Tae-Soo Jung , Eui-Ho Suh, Hyun-Seok Hwang, 2006, "Customer segmentation and strategy development based on customer lifetime value: A case study" Expert Systems with Applications", Elsevier, Volume 31, Issue 1, pp 101–107.
- [4]. Er.Arjit Gupta, Er.Ankit Gupta,Er. Amit Mishra, 2011, "Research Paper on Cluster Techniques of Data Variations", International Journal of

- Advance Technology & Engineering Research (IJATER), Vol 1 Issue 1, pp.39 - 47.
- [5]. Sankar Rajagopal, 2011 “Customer Data Clustering Using Data Mining Technique” International Journal of Database Management Systems ( IJDMS ) Vol.3, No.4.
- [6]. J. K. Pal, 2011, “Usefulness and applications of data mining in extracting information from different perspective”, Annals of Library and Information Studies. Vol 58, pp.7-16.
- [7]. K. Tulankar, M. Kshirsagar, R. Wajgi, 2012, “Clustering Telecom Customers using Emergent Self Organizing Maps for Business Profitability”, International Journal of Computer Science and Technology, Vol 3 Issue 1, pp.256 – 259.
- [8]. V.Venkateswaran, Ramakrishnan, A. Ramalingam, 2012, “E-Learning Retrieval System Through Advanced Data Mining Clustering Technique”, International Journal of Modern Engineering Research (IJMER), Vol.2, Issue.4, pp-1572-1575.
- [9]. Dharminder Kumar, Deepak Bhardwaj, 2011, “Rise of Data Mining: Current and Future Application Areas”, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, pp.256-260.