Abstract-This work helps in identifying the activities segmentation of mobile customers of different groups will be done based on some rules. Customers are segregated into groups under the categorization of network providers. There are different types of networks, but we used only four types of networks. These networks are network a, b, c and d. We can calculate the score with the help of different attributes. These attributes are- STD Calls, Local Calls, Local SMS, STD SMS, Roaming, Tariff Plan, and Data Plan. Second, Revenue maximization profiling-use of clustering for identifying maximal cluster. With the help of clustering profiling is done for maximization of revenue. This all will be done using data mining tools which will provide data statistics. In second problem Revenue maximization profiling was done by using clustering for identifying maximal cluster. Revenue maximization was done individually for each cluster.

Key Words: Data base of mobile customers, Data mining, Weka tool, K-mean clustering and Profiling

1. INTRODUCTION

The telecommunications industry was one of the first to adopt data mining technology. This is most likely because telecommunication companies routinely generate and store enormous amounts of high-quality data, have a very large customer base, and operate in a rapidly changing and highly competitive environment. Telecommunication companies utilize data mining to improve their marketing efforts, identify fraud, and better manage their telecommunication networks. However, these companies also face a number of data mining challenges due to the enormous size of their data sets, the sequential and temporal aspects of their data, and the need to predict very rare events—such as customer fraud and network failures—in real-time. The popularity of data mining in the telecommunications industry can be viewed as an extension of the use of expert systems in the telecommunications industry.

Call Detail Records (CDRs) describe customer utilization behaviour. In this paper, clustering of mobile customers is done based on the call detail records and using Self Organizing Maps. SOMs convert high dimensional data to a lower dimension representation scheme (a two-dimensional map) that can be easily visualized and understood.

With the globalization of world economy and market internationalization, the market competition of the domestic mobile communications industry has become more reasonable and increasingly fierce. The fast-growing user group, diversified operations and competition environment have put forward higher requirements for the service quality of the mobile communications industry. The competition to acquire and retain customers among mobile service providers is fierce. The key to survival in this competitive industry lies in knowing the customers better. Different people have different preferences in using mobile telecommunication services and mobile phones. Consider for example popular mobile service SMS. For a given population, majority of subscribers may be sending SMS every day whereas few may be such that they do not use SMS frequently.

Significance Of The Problem

• Segmentation of mobile customers of different groups will be done based on some rules. Customers are segregated into groups under the categorization of network providers. There are different types of networks, but we used only four types of networks. These networks are network a, b, c and d. We can calculate the score with the help of different attributes. These attributes are- STD Calls, Local Calls, Local SMS, STD SMS, Roaming, Tariff Plan, and Data Plan.

• Revenue maximization profiling-use of clustering for identifying maximal cluster.

2. RESEARCH METHODOLOGY

For solving the above two problems some research techniques and methodologies are used for obtaining the desired result. Some tools and algorithms are required for obtaining the result. Main steps under the research methodologies are:-

Review literature or research papers – first of all literatures and research papers were reviewed for getting more information about the problem and knowing which type of work was done by others on this topic and by which method.

Identify tools – then tools required for solving the problem were identified and the best tool was selected from all.

Study database attributes and data structure – attributes and structure of the database was thoroughly studied for finding out useful attributes from the mobile office.

Organize filed visits to mobile office, From there we get information about the network packs, mobile services and customer database.

Determine nature and definition of research problem and work flow of the problem for getting accurate and desired result.
Organize the database with useful attributes and populate it then perform data analysis using suitable tool e.g., WEKA in order to generate the result.

3. Conceptual Framework

Clustering technique of data mining is used to solve the problems of this work. Clustering is a method of unsupervised learning and a well known technique for statistical data analysis. Clustering is a division of data into different groups. Data are grouped into clusters in such a way that data of the same group are similar and those in other groups are dissimilar [7]. 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 [8]. Clustering aims is the objects in a group should be similar (or related) to one another and different from (or unrelated to) the objects in other groups. The greater the similarity within a group and the greater the difference between groups, the better the clustering [9].

K-Means [10] is one of the simplest unsupervised non-hierarchical learning methods among all partitioning based clustering methods. It classifies [11] a given set of n data objects in k clusters, where k is the number of desired clusters and it is required in advance.

There are two main parts:
1. Profiling Of Customers
   • Segmentation of mobile customer improving profitability using mining techniques.

The clusters obtained from ESOM need to be analyzed for the values depicted for the 16 features chosen. Component planes for each individual component could be drawn. The mean value for each of the 16 features for each cluster would be computed. Further derived measures for each cluster such as total number of Voice calls as a percentage of total outgoing calls, total number of SMS calls as a percentage of total outgoing calls etc. can be computed. Distinguishable marketing strategies can be designed for each cluster depending on the value depicted by the cluster for each of the derived measure. Following are some examples: Design suitable short message price policy for customers of cluster which has highest short messaging index value. Marketing managements need to encourage customers having low consuming ability to use more mobile service. It can be known which group of mobile customers often travels out and suitable roaming policy could be designed for customers belonging to this cluster.

![Fig. Profiling of customers](image)

Score Calculation
Score calculation is the first problem of my work. For this one database was used. Calculation of score was done using different types of packages.

Calculating Total Score –
Total score was calculated by summing the STD calls local calls local SMS, STD SMS, roaming, tariff plan, and data plan. Then we get total score for every entry package for mobile customer. Then total score is computed simply getting sum of ALL PACKAGES.

\[
\text{TOTAL SCORE} = (\text{LOCAL CALL}) + (\text{STD CALL}) + (\text{LOCALSMS}) + (\text{STD SMS}) + (\text{ROAMING}) + (\text{TARIFF PLAN}) + (\text{DATA PLAN})
\]

![Fig. Attributes of all networks](image)

Note: all the data has simulated. It is not correct information of customers.
2. Revenue maximization profiling-use of clustering for identifying maximal cluster

In second problem Revenue maximization profiling was done by using clustering for identifying maximal cluster. Revenue has been calculated manually and the formula is:-

\[
\text{Revenue of Network Provider}(S) = \sum_{k=0}^{n} \text{Total}
\]

5. RESULTS AND DISCUSSION

Profiling of customers

In first problem that is profiling of customers. In this segmentation of mobile customer, customers are segregated into groups under the categorization of network provider, was done using different datasets. From these datasets we can calculate score .then according to the score we can classify the customer as Premium and Non-premium.

There are FOUR types of clusters- cluster0, cluster1,cluster2,cluster3
Cluster 0 -> NETWORK C
Cluster 1 -> NETWORK D
Cluster 2 -> NETWORK A
Cluster 3 -> NETWORK B

1.1. CLUSTER 0
- Cluster 0 HAVING NON PREIMUM CUSTOMER TYPE
Total Entries = 66 (26%)

1.2. CLUSTER 1
- Cluster 1 HAVING NON PREIMUM CUSTOMER TYPE
Total Entries = 61 (24%)

1.3. CLUSTER 2
- Cluster 0 HAVING NON PREIMUM CUSTOMER TYPE
Total Entries = 56 (22%)

1.4. CLUSTER 3
- Cluster 0 HAVING NON PREIMUM CUSTOMER TYPE
Total Entries = 72 (28%)

Revenue maximization:-

<table>
<thead>
<tr>
<th>Revenue Network Provider</th>
<th>Cluster No.</th>
<th>Premium (Rs.)</th>
<th>Non Premium (Rs.)</th>
<th>Total (Rs.)</th>
<th>Max. Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>17,790</td>
<td>28,935.5</td>
<td>46,726</td>
<td>Non Premium</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>38,516</td>
<td>31,117.5</td>
<td>69,634</td>
<td>Premium</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>32,763</td>
<td>30,060.5</td>
<td>62,824</td>
<td>Premium</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>21,709</td>
<td>31,839</td>
<td>53,548</td>
<td>Non Premium</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1,08,558</td>
<td>1,21,752.5</td>
<td>2,30,310</td>
<td></td>
</tr>
</tbody>
</table>

Dominant B D B
A, B, C & D are different –different type of network provider. There are two categories of customers in different-different networks Premium and Non-premium. According to revenue maximization A and D are Non premium type customers and B and C are premium type customers. In the premium and non premium type-B and D both are dominant respectively. According to usage pattern total of all customer-B is the dominant.

6. Conclusion & Future Scope
Segmentation of mobile customers of different groups will be done based on some rules. Customers are segregated into groups under the categorization of network providers. There are different types of networks, but we used only four types of networks. These networks are network a, b, c and d. We can calculate the score with the help of different- different attributes. These attribute are- STD Calls, Local Calls, STD SMS, Roaming, Tariff Plan, and Data Plan. Revenue maximization profiling-use of clustering for identifying maximal cluster.

Future Scope
There is always a scope of improvement in any research and so is with this work also. This work used simulated database and not real one, so it is a kind of algorithm using which real world problem can be solved in future if real world data becomes available. In future we will match the entries between all four databases using template matching or any other matching algorithm; this will reduce the human work. We will also use different algorithms for clustering and compare the result among them.

Acknowledgement
Authors would like to thank to their head Dr. Rajan Vohra, HOD of CSE & I.T department, PDMCE, Bahadurgarh, for his valuable support and help.

References