A new approach to revenue estimation in Telecommunication Industry using Linear Model

Narina Thakur[1], Shweta Gupta[2] and Dr. Abhay Bansal[3]

[1]Phd Research Scholar, Department of Information Technology, Amity University Noida India and Associate professor in Department of Computer Science Engineering , Bharati Vidyapeeth College Of Engineering New Delhi, India

[2] UG Scholar Department of Information Technology, Bharati Vidyapeeth College Of Engineering New Delhi, India.

[3]Head of Department, Department of Information Technology, ASET , Amity University Noida India.

Abstract

Planning prediction based on quantitative analysis is becoming need of the day for each and every Industrial sector. The cause can be the advancement of new technologies, hence forth competition in the Industry. In the Business sector there are many prominent constraints, such as the intake of products, the quality of services offered, the kind of services and quantity of customers, performance of product, investment and hence forth revenue generation.

This paper verify and analysis the effectiveness of linear model for estimating revenues in Telecom using Binomial family, but it can be successfully implemented for any application.

Keywords: *Linear Model, Binomial family, Data Mining, Clustering, CLARA, R tool, PAM, Prediction, LD50.*

1. Introduction

Internet is a huge assert of almost every kind of information. It can be for an Individual, Enterprise businesses, government agencies and policies. With the rapid development in mobile telephone network, video and Internet technologies, huge competitive pressure is generated on the various companies and companies are looking for benchmarks for enhancing their performance or improve business plans for generating more revenues against competitor's to remain an lead player in the market.

Huge data is generated by Telecommunication companies. These data include call detail data, which describes the calls that traverse the telecommunication networks, network data, which describes the state of the hardware and software components in the network, and customer data, which describes the telecommunication customers [1]. Data mining can be used to extract / discover knowledge concealed within Telecommunication data sets. Various data mining tools are to detect telecommunication fraud, improve marketing effectiveness, and identify network faults. In Telecom industry churn prediction and management is becoming a great concern to the mobile operators. Mobile operators wish to satisfy their subscribers need in order to retain them.

So far restrictions is only on voice call rates like per second billing; telecom operators are in an overdrive to reduce tariffs on roaming and SMS charges as well. The main idea behind this for the various telecom companies is to reduce customer churn and maintain their market share. The plans which the various telecom operators can propose can be: reduced the roaming charges by almost 60%, [2] extended per second billing mechanism to roaming. Telecom companies' contention policy also revolutionizes the competition among various companies. Telecom companies' should be paid for allowing subscribers to shift from one network to another using mobile number portability.

2. Data mining technique

Data mining, the extraction of hidden values from our data warehouse, used to predict future information from large databases. It enhances Business model and change policies based on patterns. In recent years Data is enormous and analysis of the data makes sense. Data mining acts as a powerful technology which can answer various business queries helps the organization to grow. Various mining tools are available to forecast future trends and behaviors, allowing businesses to make practical, knowledge-driven decisions.

Predictive Data Mining is an upcoming area which is used to propose a model that can be used to predict particular response of interest. For example, a credit card company may be interested in identify fraudulent transactions or an

Organization may be interested in recognizing segments of privileged customers. Data Reduction is another possible objective for data mining which is summative or to merge the information in very large data sets which is useful and the data chunks are manageable.

Data Mining for Predictions:

Table1. Prediction Table

	Application Area	Yesterd ay	Today	Tomor row
Static information and current plans	Demograph ic data and marketing plans	Known	Known	Known
Dynamic information	customer transactions	Known	Known	Target

Scope of Data Mining Techniques:

Data mining is similarities between searching for treasured business facts/figures in a large database. Data mining processes require examining massive amount of material, or intelligently examining it .Data Ming technology can generate new business opportunities by examining massive transactional database. Data mining technology provides various capabilities as:

- Automated prediction of trends and behaviors.
- Automated discovery of previously unknown patterns. [5]

The most commonly used Data mining techniques are:

- Artificial neural networks: Neural Network is a well-established approach for modeling data.[6,7] It is a Non-linear predictive models which learn through training and resembles biological neural networks.
- Decision trees: It is Tree-shaped structures that represent sets of decisions which generate rules for the classification of a dataset.
- Genetic algorithms: GA is the optimization techniques that use genetic combination, mutation, and natural selection processes.
- Nearest neighbor method: It is a technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset.
- Rule induction: if-then rules extraction from data based on significance.

3. Data mining process

- **Pre-Processing the Data:** Data mining and data analysis includes identifying, gathering, and processing the data that will be analyzed. The first task is to remove from the database those attributes/ values which are irrelevant followed by Data preprocessing process involve Cleansing, collapsing. We first identify what the investigation is intended to discover and type of data that will be useful is a very complicated task.

- **Finding Search Models:** This process can be an automated data analysis or deliberating pattern-based searching, finding and finalizing those models can be a very complex and difficult task. There are several ways to come up with the patterns on which a model is based. Models can be found using data-mining analysis. The data Analysis is a “bottom-up” approach, which predicts a model in data viz. it starts with the data and looks for outlier or patterns that indicate certain behavior.

- **Decision-making:** The concluding stage in data-mining and analysis process includes searching, interpreting the results and making decisions/ policies about the result usage. For example, a retailer might apply data-mining models to predict the buying behavior of shopper’s based on his historic purchases, automatic recommendations are sent to the shopper, without the intervention of an employee.

4. Linear model

Linear Model is the simplest model; we try to predict the outcome of one variable from one or more predictor variables. [3]

$Yi = (b0 + b1Xi) + ei$

Yi = outcome we want to predict

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 3, May 2012
ISSN (Online): 1694-0814
www.IJCSI.org

b0 = intercept of the regression line regression
b1 = slope of the regression line coefficients.
Xi = Score of subjection the predictor variable
ei = residual term, error

Binomial family

Consider the Greek island of Kalythos problem which gives the prediction that the male inhabitants of age 43.663 years and 43.601 years are more prone to congenital eye disease. This problem is similar to silver 1960 problem and AI problem [4]. The samples of islander males across various ages were tested for blindness and data is recorded and the solution to the above problem is estimated for LD50 that is the age at which the chance of blindness for a male inhabitant is 50%.

5. Telecom research variables

Table2. Telecom Research variables and its explanation Implementation

Sr. No	Input variables	Description
1.	OG_MOU(in Mins)	Outgoing Minutes of Usage i.e. number of minutes for which subscriber made outgoing calls (excluding roaming OG_MOU).
2.	OG_AIRTIME	Outgoing Airtime revenue i.e. revenue generate from OG_MOU or balance deduction from subscriber's account.
3.	R_OG_MOU(in Mins)	Outgoing minutes of usage while roaming out of its home location/circle.
4.	R_OG_AIRTIME	Airtime revenue generated while subscriber is on roaming.
5.	NETARPU	In Net ARPU we deduct all payouts charges such as (Access Deficit charges ADC, Interconnectivity charges e.t.c) from Gross ARPU which will give Net ARPU per subscriber.

5. Implementation

Implementation uses ‘R’ tool, which is a Programming Environment for Data Analysis and Graphics. R is an integrated suite of software facilities for data manipulation, calculation and graphical display. Clustering analysis for customer segmentation and visualization is done using R. Telecom data Implementation steps are as explained in Fig1 and explanation is as given below:

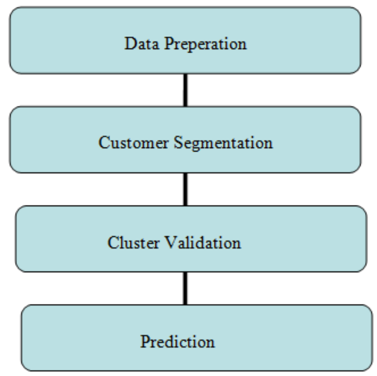

Fig1. Implementation steps

1. **Customer Segmentation**: The technique used for customer segmentation is clustering. A cluster can be defined as a collection of objects which are “similar” between them and “dissimilar” to the objects belonging to other clusters. The similarity criterion that was used in this case is distance: two or more objects belong to the same cluster if they are “close” according to a given distance (in this case geometrical distance). This is called distance-based clustering.

2. **Clustering Algorithm**: We are using here CLARA clustering algorithm, which is clustering large Application. Compared to PAM, CLARA deals with large data set. Instead of finding medoids for the entire data set, CLARA draws a small sample from the data set and applies the PAM algorithm to generate an optimal set of medoids for the sample. The run time complexity is O(n) instead of O(n2) as in PAM.

3. **Cluster validation:** Cluster validation refers to the problem whether a found partition is correct and how to measure the correctness of a partition. A clustering algorithm is designed to parameterize clusters in a way that it gives the best fit. A variety of measures aimed at validating the results of a clustering analysis

and determining which clustering algorithm performs the best for a particular problem statement.

Prediction

Customers are offered plans based on revenue generated based on Outgoing minutes of usage normal and revenue generated based on Outgoing minutes of usage based on roaming. We can offer plans for either/both based on outgoings minutes of usage (Normal/roaming) or on NETARPU.

Before predicting customer category we apply LD50 Binomial model to the data

Table3.Results obtained from Telecom Data set using R-tools and CLARA clustering method.

OG_MOU(in Mins.)	160.92	154.79	148.88	143.26	137.83
R_OG_MOU(in Mins.),	170.97	133.83	103.57	95.34	79.13
NETARPU	230.52	211.61	188.99	179.00	170.13
Sample size	50	50	50	50	50
No. Custo-mers	6	6	11	12	15

```
> xsub<-x[1:50, ]
> clarax<-clara(xsub,5)
> print(clarax)
Call:    clara(x = xsub, k = 5)
Medoids:
       V1    V2     V3    V4     V5
4  160.92 87.58 170.97 87.58 230.52
9  154.79 81.00 133.83 81.00 211.61
18 148.88 75.54 103.57 75.54 188.99
29 143.26 71.53  95.34 71.53 179.00
43 137.83 66.67  79.13 66.67 170.13
Objective function:7.348205
Clustering vector: Named int [1:50] 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 3 ..
 - attr(*, "names")= chr [1:50] "1" "2" "3" "4" "5" "6" "7" ...
Cluster sizes:            6 6 11 12 15
```

Fig2. Results for CLARA for k=5 and n=50

Here V1, V2, V3, V4 and V5 are mediods for OG_MOU (in Mins), OG_AIRTIME, R_OG_MOU (in Mins.), R_OG_AIRTIME, and NETARPU.

Using Binomial and LD50 model of Rtool the LD50 intercepts are calculated as:

For OG_MOU(in Mins) ld50 Intercepts is 109.2282 and 113.7801.

For R_OG_MOU(in Mins) ld50 Intercepts is 117.6881 and 121.1987.
For NETARPU ld50 Intercepts is 98.2455 and 110.4331.

Based on the OG_MOU, R_OG_MOU and NETARPU we offer three different plans to our Telecom customers, which is as shown in Table3.

6. EXPERIMENTAL RESULT

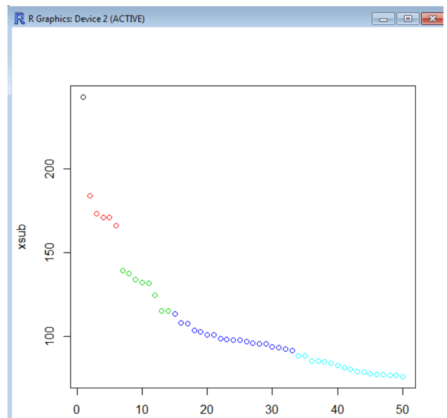

Fig3(a). Visualisation for CLARA Clustering of Telecom dataset for k=5 for sample size 50.

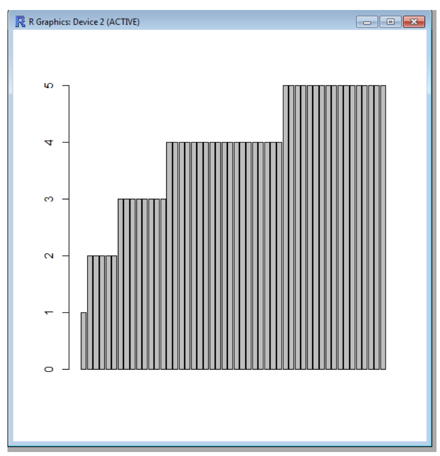

Fig3(b)Bar plot for CLARA Clustering of telecom dataset

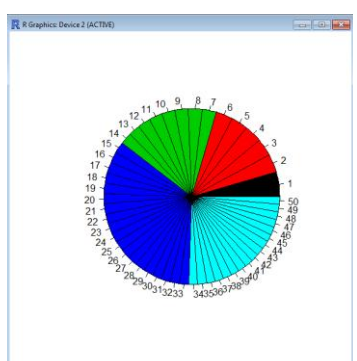

Fig3(c) Pie Char for CLARA Clustering of telecom dataset

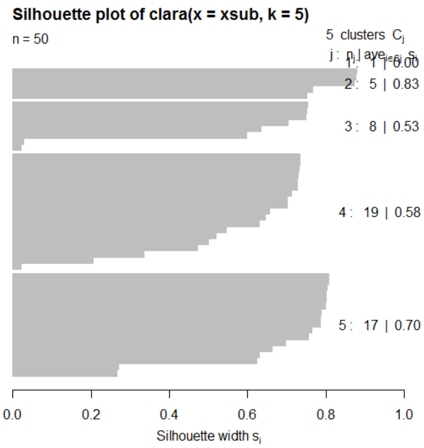

Fig3(d)Silhoutte plot of clara for k=5 and n=50 Average Silhoutte width is 0.63.

Table4. Telecom Scheme offered

Plan	Research condition	Scheme Offered
Plan-A	OG_MOU(in Mins.) >= 111.5	Telecom service area: All circles, Monthly rental (in Rs.): 199, for same Mobile operator: 1p, Mobile operator to landline: 1.1p, Mobile operator to different Mobile operator: 1.1p.
Plan-B	R_OG_MOU (in Mins.) >= 119.4	For Different mobile operators: 0.7, STD to local: 0.6, STD to different local Mobile operator : 0.8, Incoming calls:0.5 Free benefits
Plan-C	NETARPU(in Rs.) >= 104.3393	Free local second : 10000, Free Local SMS: 0.5

7. Conclusion

In this paper we have introduced the idea for estimating revenue in telecom sector by proposing various plans to the customers based on their Outgoing minutes of usage in normal mode/ in roaming mode and Net Airtime minutes of usage.

We have shown how to compute an approximation of OG_MOU, R_OG_MOU, NETARPU. Here we have described and used LD50 Binomial family Linear model and Clara Clustering in the field of Telecom sector for predicting plan and client segmentation. CLARA run time complexity is O(n) instead of O(n2) as in PAM.

References

[1] The Analysis of the Telecommunications sector by the means of Techniques Adrian COSTEA Journal of Applied Quantitative Method Interdisciplinary – New Approaches and Perspectives in the Use of Quantitative Methods no 2 vol 1.doc Winter 2006.
[2] http://indiatelecomnews.com/?p=931
[3] http://en.wikipedia.org/wiki/Linear_model
[4] http://www.math.montana.edu/Rweb/Rnotes/node87.html
[5] www.thearling.com/index.htm
[6] Artificial Neural Networks Ajith Abraham Oklahoma State University, Stillwater, OK, USA, Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn. 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8
[7] G. Grudnitski and A. Quang “Important factors in neural networks’ forecast of gold futures prices,” in Proc. Of the 4th International Conference of Neural networks in the Capital Market, 1996,pp-163-176.

NARINA THAKUR is presently an Associate Professor of Computer Science Department in Bharati Vidyapeeth’s College of Engineering, New Delhi (GGSIPU). She is a B. Tech gold medalist and pursued M.Tech from Punjab Technical University with exceptional career Record and pursuing PHD from Amity University Noida, Apart from teaching, she has also authored many books in the areas of Algorithms and programming and Operating system. She is an active member of Research and Development team in Bharati Vidyapeeth and working closely in the areas of Algorithms, Information Retrieval, Data mining, and Embedded Systems. She has also authored many papers, published in IEEE International Conferences, ACM SIGARCH and International Journals.

Shweta Gupta is a final year undergraduate student, in Bachelors of Technology, Information Technology, Bharati Vidyapeeth’s College of Engineering; New Delhi affiliated to GGS Indraprastha University (New Delhi). She is the founder of Indian Society of Technical Education (ISTE) student chapter in the college. She is also a student member of Research and Development cell of the college. Apart from being a top ranking student, she has been actively involved in the following research areas: Data Mining, Text Retrieval and Analysis, Clustering, Artificial Intelligence, Natural Language Processing, Information Extraction.

Dr. Abhay Bansal is BE (CS), ME (IT), MBA and PhD. He is Professor & HOD (IT) with Amity School of Engineering & Technology, Amity University Noida. With over 17 years of Industry and Academic Experience. Dr. Bansal is guiding 8 students for Ph.D and also member of DRC of Amity University, Noida. Dr Bansal is also member of examination committee of several of university. Dr. Bansal is Fellow, The Institution of Engineering and Technology (U.K), Sr. Member International Association of Computer Science and Information Technology, Member International Association of Engineers, External Member of RDC of Nothern India Textile & Research Association (Ministry of Textile, Govt. of India) and Sr. Life Member of Computer Society of India.