

Deriving Insights and Analysis for Campaign Management

Shruti Mahajan¹, Devika Antarkar², Ryan Roy³, Manish Nagare⁴, Rakhi Kalantri⁵, Shagufta R⁶

Computer Engineering, Mumbai University, Navi Mumbai, Maharashtra, India-400703

Email: rakhi_kalantri@rediffmail.com

Abstract— There has been a rapid increase in the growth of telecommunication industry. Customer relationship has become a factor of competitive advantage. Over the past few decades, companies have accumulated tons of customer data. The analysis of this data can improve the chances of survival of that company in the highly competitive market. Campaigns are an efficient way of advertising the schemes, offers and services offered by a telecommunication industry. The campaigns are custom-made according to customer usage pattern and behavior. Accordingly, aggregation and segmentation is performed. The campaigns are an important part of marketing products and services. Campaigns are sent through various channels (SMS, Email etc). The response generated through each channel will be analyzed for further campaigns to be sent. This demand of continuous analysis of historical as well as current data for enhanced campaign management can be achieved with the use of various data mining technologies. R is a powerful tool which provides easy analysis of large amount of data. It has various packages which has different inbuilt functionalities which can easily produce the output of the desired mathematical problem in a few seconds. For this reason, R is a widely used statistical programming language used by analysts worldwide. This research paper aims at explaining how different data mining methods are implemented with R for Campaign management.

Keywords— Campaign, R Tool, Telecommunication, Data mining, Aggregation, Segmentation.

I. INTRODUCTION

As the technology advances, there has been a significant change in the way companies advertise their products. Marketing strategies have become smarter and less time-consuming. Reliance Jio, being a telecommunications company receives and maintains a large amount of information, based on which they have to alter their marketing frequently.

There are many sources through which the Analytics team receives data. These sources includes various Jio apps such as JioChat, JioBeats, JioDrive, JioOnDemand, MDM(Master data management) which stores the static customer information such as Name, Address, etc. and is updated every time a new customer joins the service. SAP CC and SAP HANA are the two sources which store the charging information. In addition to this, periodic surveys are conducted to gauge the response of a customer to a particular scheme or offer.

There are several key parameters (KPI'S) defined in SAP CC which are the attributes on which queries are processed. Based on these KPI's the marketing team can keep track of the data usage done by the customers and can predict the general pattern of their future usage. The marketing strategies are then developed to increase the company revenue by pitching more schemes to active users and also to the passive user from leaving the service provider. This is how the basic segmentation of data is performed. These schemes are then pitched to customers via different channels such as SMS, Email, etc[1].

The data organization and storage strategies used by an organization should aim to retrieve each customer's

required information as quickly and efficiently as possible. Achieving this will ensure that there is no delay in processing of this information by the analyst team for the generation of suitable and customized campaigns for every customer.

This process of retrieving and analyzing data should be optimized for time efficiency but should also be straightforward enough so that it can be understood by different analysts.

II. LITERATURE SURVEY

A. Overview

The market needs change every moment. The supply and demand of products and services vary proportional to each other. There are various aspects of customer-relationship management, business intelligence, intervention analysis and various technologies and tools well-suited for extracting and segmenting data to meet the market and the customer needs. Business intelligence encompasses various objectives to be met in order to achieve various Operational efficiency, Optimizing internal business processes, etc. Business intelligence incorporate various advanced data analytics, such as data mining, predictive data analytics, intervention analysis, statistical analysis, Big data analytics, etc. Additionally, Hadoop systems are increasingly used within Business Intelligence architecture as repositories. Raw Data is integrated and consolidated into different sources using data integration and data quality tools to ensure that users are analyzing accurate and consistent information. Organizations are looking to optimize potential social interactions and increase familiarity with customers by

developing relationships with various stakeholders through social media platforms. Customer Relationship Management including, but not limited to, social CRM strategies and technologies, creation and management of customer's networks, customer dynamics, social media analytics, customer intelligence, word of mouth advertising, etc.

"CRM includes a particular dimension, relying on optimizing customer value. Generally, CRM is based on features able to deliver real-time analysis, both quantitative and qualitative, instead of the standard reports issued by similar applications. Usually, the most frequent reports issued through CRM relate to customer ranking (including demographical details), their critical perspective or service/product profit. CRM is able to offer strategic information for an improved planning of resources, as consequence the company is aware of the reliable strategy to implement, as well as the possibility of re-evaluating or re-structuring the unsatisfactory parameters.

Generally, ERP vendors as Oracle or SAP attempt to expand their products on back-office level, while their presence on front-office is already significant. CRM applications, created by companies as Cisco, HP or IBM focusing on hardware, software and IT infrastructure tend to supply both front-office and back-office needs, challenging the position of ERPs over the next period." [1]

Various technologies are engaged into this system such as MDM, SAP HANA, SAP CC, etc. are majorly used as sources to extract information for performing segmentation and aggregation for the given requirement of campaign management. SAP CC is a cost converging system which stores all user-related billing information based on usage (data, voice, SMS, etc). This helps to analyse the consumer usage, their behaviour and accordingly forward the given campaign for increasing consumer base and revenue generation efficiency.

B. Campaign Management Process

Figure 1 illustrates the closed loop Campaign Management Process. The various steps involved in Campaign Management are:

The data is collected through various sources such as SAP CC, Jio Apps etc which is then loaded in the database. The

next step is to perform segmentation and aggregation operations on the set of customers, which is done by the data analysts. This is done so that campaigns are sent to appropriate group of customers which have a higher likelihood of responding to them. The campaign management team is then responsible for sending these campaigns to customers via different channels such as SMS, Email, Direct Call etc. [2]

The next step is to obtain feedback from customers by sending surveys or different types of feedback form. Depending on this feedback, alterations can be made to campaigns to improve customization (i.e. the suitability of the campaign to the customer) and to also increase the chances of the customers of responding to campaigns. These steps are essential for any telecommunication industry to generate revenue and also to survive in the market.

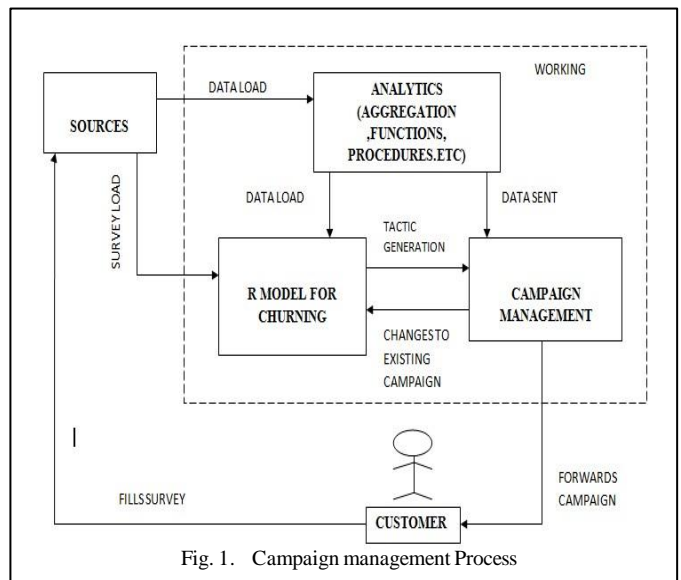


Fig. 1. Campaign management Process

III. TOOLS AND REQUIREMENTS

A. Software Requirements

The software requirements for the project are as follows:

1. Greenplum Database

The system consists of a master node, standby master node, and segment nodes. All of the data resides on the segment nodes and the catalog information is stored in the master nodes.

2 R Studio

RStudio is a free and open-source integrated development environment (IDE) for R, a programming language for statistical computing and graphics. RStudio is a free and open-source integrated development environment (IDE) for R, a programming language for statistical computing and graphics

3. SAP CC

Convergent charging and billing (also known as convergent charging, converged charging and convergent billing), is a solution in the telecommunication industry that enables common management of all users and all services for operators.

4.RShiny

The RShiny package is open source and helps in effective visualisation through using Rmarkdown package which helps create an application and an interactive interface for analysis.

B. Hardware Requirements

This project deal with analysis of data derived from existing sources prevalent and used by the organization. There are no direct hardware requirements for the project, however since we are dealing with a large volume of data appropriate data warehousing techniques and storage facilities are utilized. Various data centers manage multiple databases for staging and deployment.

Database version: Greenplum (10 TB plus)

C. Analysis Tool Used- R''

The R programming language is an open source scripting language for predictive analytics and data visualization.

It includes functions that support linear modeling, non-linear modeling, classical statistics, classifications, clustering and more. It has remained popular in academic settings due to its robust features and the fact that it is free to download in source code form under the terms of the Free Software Foundation's GNU general public license. It compiles and runs on UNIX platforms and other systems including Linux, Windows and macOS.

The R language programming environment is built around a standard command-line interface. Users leverage this to read data and load it to the workspace, specify commands and receive results. Commands can be anything from simple mathematical operators, including +, -, * and /, to more complicated functions that perform linear regressions and other advanced calculations.

Users can also write their own functions. The environment allows users to combine individual operations, such as joining separate data files into a single document, pulling out a single variable and running a regression on the resulting data set, into a single function that can be used over and over.

The appeal of the R language has gradually spread out of academia into business settings, as many data analysts who trained on R in college prefer to continue using it rather than pick up a new tool with which they are inexperienced.

IV. ALGORITHM

Regression is a statistical measure used in finance, investing and other disciplines that attempts to determine the strength of the relationship between one dependent variable (usually denoted by Y) and a series of other changing variables (known as independent variables) Regression analysis is an important tool for modelling and analyzing data. Here, we fit a curve / line to the data points, in such a manner that the differences between the distances of data points from the curve or line is minimized.

There are various forms of regressions, which can be performed. Each form has its own importance and a specific condition where they are best suited to apply.

A. Linear Regression

It is one of the most widely known modeling technique especially used for predictive modeling. In this technique, the dependent variable is continuous, independent variable(s) can be continuous or discrete, and nature of regression line is linear.

Linear Regression establishes a relationship between dependent variable (Y) and one or more independent variables (X) using a best fit straight line (also known as regression line).

It is represented by an equation $Y=a+b*X + e$, where a is intercept, b is slope of the line and e is error term. This equation can be used to predict the value of target variable based on given predictor variable.

B. Logistic Regression

Logistic regression is widely used to model the outcomes of a categorical dependent variable. For categorical variables, it is inappropriate to use linear regression because the response values are not measured on a ratio scale and the error terms are not normally distributed.

Here the value of Y ranges from 0 to 1 and is represented by the following equation:

$odds= p/ (1-p) = \text{probability of event occurrence} / \text{probability of not event occurrence}$

$\ln(odds) = \ln(p/(1-p))$

$\text{logit}(p) = \ln(p/(1-p)) = b_0+b_1X_1+b_2X_2+b_3X_3...+b_kX_k$

V. IMPLEMENTATION

The project has been developed using R programming. So, the implementation part is explained below:

A. GUI Implementation

The GUI has been designed using the R Shiny packages in R tool. The GUI that we designed is as shown below:

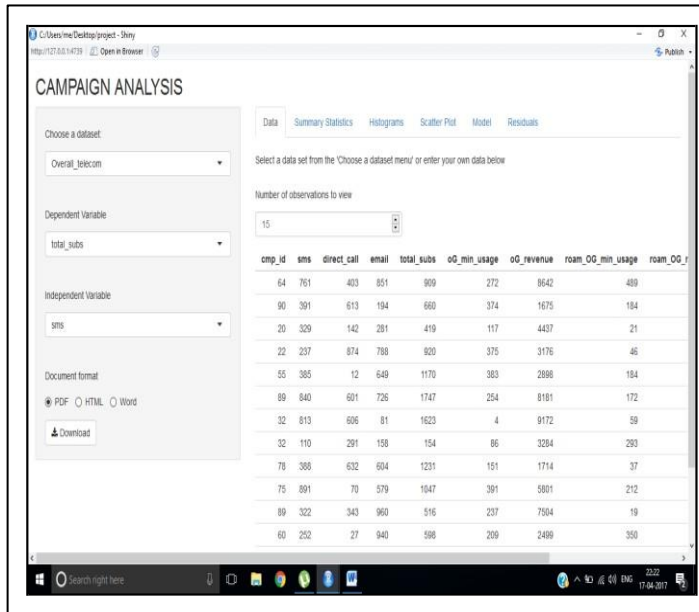


Fig. 2. Graphical User Interface

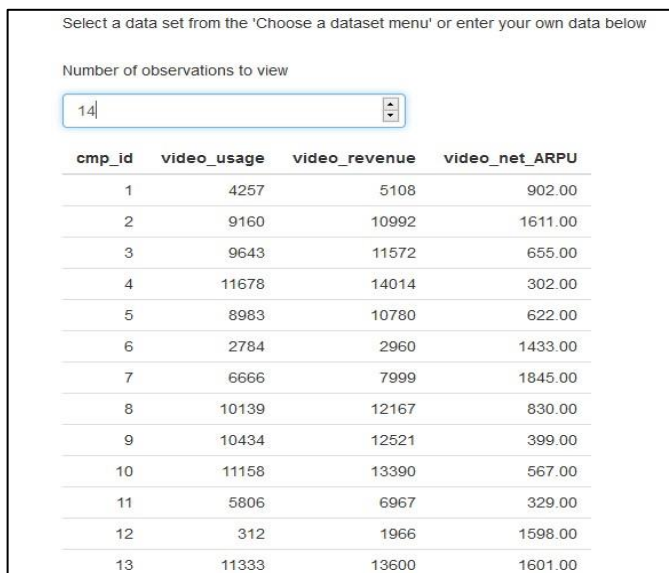


Fig. 3. Dataset chosen for analysis



Fig. 4. Linear model of dataset

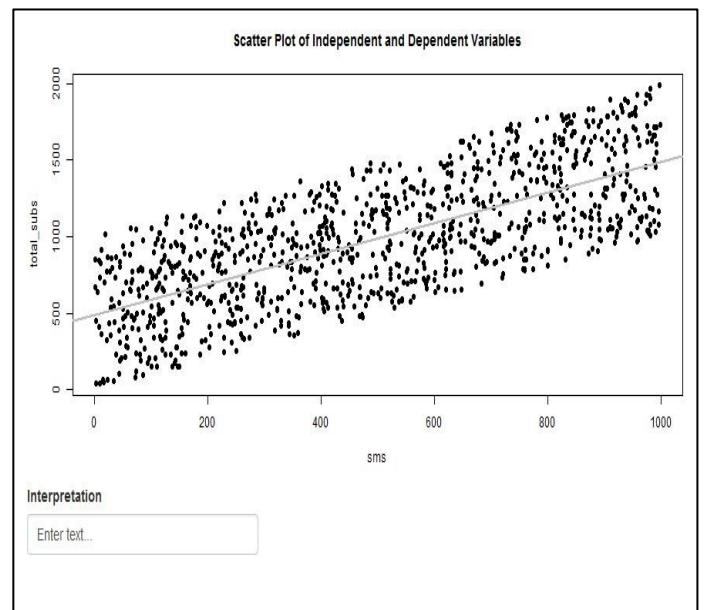


Fig. 5. Scatterplot

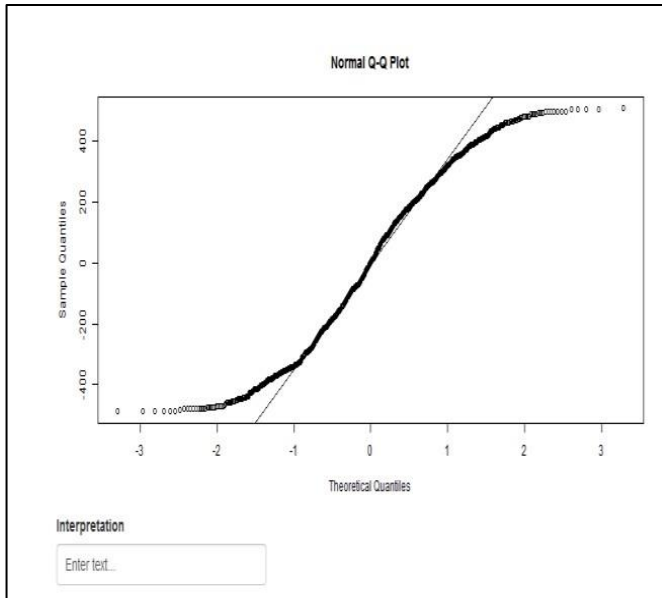


Fig. 6.: Normal Q-Q Plot

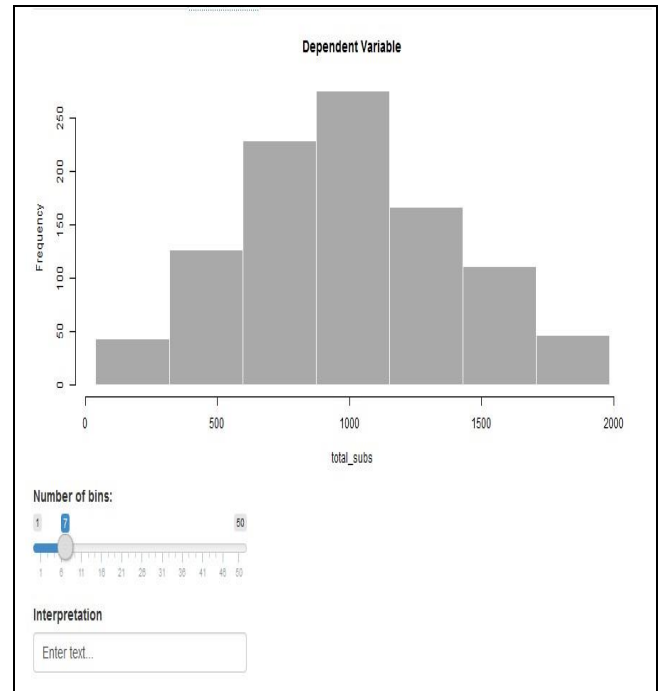


Fig.7: Histogram

The various components used for making GUI are different panels, tabs, Pushbuttons, dropdown lists, text boxes.

B. Browsing through options and data representation

In this, we have created functions that allow you to choose from different options for datasets, dependent variable, independent variable and format of document required.

The data is represented in a 2D format of rows and columns.

C. Applying linear regression

Linear regression is applied to predict scores on one variable with scores on a second variable. It is approach for modelling the relationship between a scalar dependent variable and one or more explanatory variable or independent variable.

D. Graphical representation of results:

The result of our project, described in this paper is a graphical representation of the relationship between the dependent variable and the independent variable.

Here, the line represents the ideal relationship variables and the dots are the data points deviating from the expected results.

The normal Q-Q plot shows how the data points are expected to behave in the ideal situation. This is shown in the residuals tab where we also represent their frequency.

One of the tabs also provide Histogram representation of the dataset. It's a diagram consisting of rectangles whose area is proportional to the frequency of a variable and whose width equal to the class interval.

Evaluation, sequence discovery are the techniques used in descriptive modeling to uncover relationships and trends in data. When the economy is challenging and the business landscapes are changing faster than ever, it is critical for organizations to focus on their important business issues. Thus, making the application useful to understand the nature of these challenges

VI. CONCLUSION

Business Analytics have become vital for the growth and development of the companies. This tool can be enhanced depending on the business requirements to accommodate different types of campaigns. Provide further automation by integrating it with the business module. Classification can be implemented for customers with similar usage patterns or according to the campaign rotations. Personalization can be done using detailed prediction for particular categories of customers. The data gathered by the company over the years can be used in data mining strategies to provide an insight in to the trends and patterns which can be extremely useful while analyzing data to send campaigns.

ACKNOWLEDGMENT

Authors would like to thank our guide, Rakhi Kalantri and Vijay Tiwari for their constant support and guidance and

supplying us with necessary information and data to carry out analysis.

REFERENCES

- [1] Michael Carter, Jerry Fiala, Oved Hernandez, Mark Chase Mighell, Jordan Sacks, Colin Tucker, QuanquanGu, William Scherer, "Advertising.com pre-install app campaign," *2016 IEEE Systems and Information Engineering Design Symposium (SIEDS)*, Charlottesville, VA, 2016, pp. 84-88.
- [2] A. Scharmann Anne Scharmann, Ryan Crawford, Katherine Fetscher, Ryan King, Yi Lu, Kevin Ngo, William Scherer, "Advertising.com: mobile optimization and predictive segments," *Systems and Information Engineering Design Symposium (SIEDS), 2015*, Charlottesville, VA, 2015, pp. 80-84.
- [3] Lian Yan, M. Fassino and P. Baldasare, "Predicting customer behavior via calling links," *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, Montreal, Que., 2005, pp. 2555-2560 vol. 4.
- [4] M. Goul, S. Balkan and D. Dolk, "Predictive Analytics Driven Campaign Management Support Systems," *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, Kauai, HI, 2015, pp. 4782-4791.
- [5] D. Kridel, D. Dolk and D. Castillo, "Adaptive Modeling for Real Time Analytics: The Case of "Big Data" in Mobile Advertising," *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, Kauai, HI, 2015, pp. 887-896.
- [6] D. Kridel and D. Dolk, "Latticing and Device-Histories: Dynamic Customer Profiling for Mobile Advertising Campaigns," *2016 49th Hawaii International Conference on System Sciences (HICSS)*, Koloa, HI, 2016, pp. 1102-1111.
- [7] Wayne W. Eckerson, "PREDICTIVE ANALYTICS :Extending the Value of Your Data Warehousing Investment," *first quarter 2007, TDWIBest practices report*
- [8] Mike McGuirk, "Customer Segmentation and Predictive Model