Survey on CRM Analytics in Telecommunication Industry

Lewlisa Saha[#1], Laxman Sahoo[*2], Sanjib Kumar Routray[#3]

[#1]*M.Tech Scholar, School of Computer Engineering, KIIT University Bhubaneswar, Odissa, India- 751024*

[#2]*Professor, School of Computer Engineering, KIIT University Bhubaneswar, Odissa, India- 751024*

[#3]*Assisstant General Manager, BSNL, Bhubaneswar, Odissa, India- 751012*

Abstract— *The goal of data analytics is uncovering fruitful information and decision making from very large data sets in respect to various applications, which can redefine and improve customer relationships. With the increasing need of customizing every product and services the concept of Customer Relationship Management has been introduced. CRM is practiced in business by detecting patterns in customer data. Telecommunication sector is most well known and important to be dealing with customer interests than anyone else since they are well aware of their customers and can efficiently keep record of their customers' actions. As the colossal volume of data produced by telecommunication companies cannot be evaluated manually, varied data analytics approaches are needed to be applied. Data analytics assists a business grasp its customers in a much superior way. In this paper we will make a survey on the concept of CRM and the use of data analytics in it.*

Keywords — *Data Science, Data Analytics, CRM, Clustering, Pattern Recognition, Pattern Proximity Measure, Kohonen Clustering, Spectral Clustering, K-means, Telecommunication Industry, Data Abstraction* .

I. INTRODUCTION (SIZE 10 & BOLD)

Data Science is a very promising field of study to fit into the need of every organization. This is an interdisciplinary field based on the scientific processes and systems to extract the knowledge or insights of data in various forms and turning the data into actions, whether it is structured or semi-structured or unstructured. This is some how similar to knowledge discovery from the set of very large database of different types. In recent years the IT sector have gotten ingrained in our society with the rise in the amount of data available in our surrounding. Among all the different task domains of Data Science, in this paper we will be mainly talking about data analytics and how we can apply it in Customer Relationship Management (CRM). CRM is an efficient process which is being used mostly by every business organizations, since they have to maintain a good relationship with their customers, as they are the sole important entity of every business management these days[1]. In this paper, we will be mainly targeting the telecommunication industry to explore how Data Analytics techniques can be applied in order to improve their overall services.

In section 1, few of the domains of Data Science have been discussed along with the differences and common factors. We will be covering the topics of Data Analytics and its outcomes, in section 2. In next section, we will be describing CRM in brief and why Data Analytics is used in CRM and why we apply it in telecommunication industry. In section 4, we will be explaining the types of data that are available in the telecommunication industry and which data are mainly concerned with our topic of discussion. In section 5, we discuss about the techniques of Data Mining those are used in Data Analytics and we will be mainly concentrating on the Data Clustering technique in order to create a marketing profile.

II. DOMAINS OF DATA SCIENCE

Data Science cover almost all industries and fields, specifically, digital analytics, search technology, marketing, fraud detecting, astronomy, energy health-care, social networks, finance, forensics, security (NSA), mobile, telecommunications, weather forecasts, and fraud detection. Data Science is a very broad discipline, hence it can be narrowed down to a number of domain expertise which overlap with Data Science.

A. Data Mining

This field mainly concerns with designing algorithms to extract new info from large pre-existing structured data[2].

B. Predictive Modeling

It is mainly a project that develop in all industries. This application mainly target at predicting the probability of an outcome.

C. Computer Science

Data Science overlaps with Computer Science in more than one ways, like: Hadoop and Map-Reduce

implementations, Algorithmic and Computational complexity to design fast, and so many more.

D. Data Engineering

It mainly includes designing, building and managing of information. It analyze and process data by developing a suitable architecture.

E. Data Analytics

It is much narrower perspective of Data Science. Data Analytics is the operations and Data Science is the strategy[3]. According Dr. Jerry A. Smith, Data analytics seeks to provide operational observations into issues that we either know we know or know we don't know.

F. Business Analytics

It is mainly similar to Data Analytics but is restricted to Business problems only.

III. What Is Data Analytics

Data Analytics is considered to be a field of Data Mining where we analyze the data by applying different scientific approaches to reach a proper conclusion from the information available[4].

There are mainly four types of Data Analytics[3]:

A. Prescriptive Analytics

It mainly unveils what actions should be taken. This analysis mainly give the rules and recommendations for the next step. For example, prescriptive analytics can be used to measure the obesity of patients and determine where to focus the treatment.

B. Predictive Analytics

It mainly predicts the most probable scenario of what might happen by identifying the past patterns. For example some company use predictive analytics for sales lead scoring and and some companies even use it for the entire sale process.

C. Diagnostic Analytics

It is used to assess the reason behind any event that has happened. For example, in case of any social media marketing campaign, the diagnostic analysis can be used to analyze the no of posts, views, fans, etc.

D. Descriptive Analytics

It is important for analyzing patterns that can give a better perception of the data. For example, existing financial data can be analyzed to predict the customer's future conduct.

IV. Customer Relationship Management (CRM)

CRM as defined by Paul Greenberg[5]: "CRM is study of fundamental nature of knowledge, reality and a business strategy, supported by the system and technology, designed to improve human interactions to that of the business environment. It is a continuing business process which demands a dynamic, ongoing plan of action for customer agreement."

Due to the customer aligned view of the market, the business need to concentrate on the needs of the customers. In order to customize their products according to the customer's liking, every organization should study their customers closely and observe their activities. CRM mainly indicate business practices taking place across the enterprise to set-up the assets, aggregate and analyze the customer profiles[6].

A. Components of CRM

Customer Relationship Management (CRM) aims to get a 'single integrated view of customers' and a 'customer centric approach'. There are three components of CRM[5]:

- Customer
- Relationship
- Management

B. Types of CRM

There are three types of CRM[1],[7]:

1) ***Operational CRM:*** commonly refers to services targeted towards the customers to solve their problem.

2) ***Collaborative CRM:*** is a very efficient method of communication since it includes feed-backs and issue reports directly from the customers.

3) ***Analytical CRM:*** takes into consideration product and service decision-making as well as pricing and new product development.

C. Steps to Achieve Expense Targets of CRM

The main expense targets of CRM are to achieve higher profits and increased sales by the means of customer satisfaction and reducing buying and delivery costs. The basic steps to accomplish the targets are[5]:

1) ***Customer Identification:*** Getting to label the customer overtime via different websites marketing medium, surveys and communications.

2) ***Customer Differentiation:*** Differentiate each and every customer depending on their individual unique demands and requirement.

3) ***Customer Interaction:*** Interact with the customers to keep track of their behavior and needs and the changes in demand over time.

4) ***Customization:*** Customize the product in such a way that the customers are treated uniquely and their each and every demands are fulfilled.

The steps of customer interaction includes four stages[5]:

- **Initiation-** the new customers are welcomed by the management.the image used in each figure is clear,
- **Integration-** the products or services are assessed by the management.
- **Intelligence-** the management checks all the previous stages to assess the customer performance to maintain a long term relationship.
- **Value Creation-** establishing an alliance with the customers to an intricate part of their operations.

D. Link Between Data Analytics and CRM

Data Analytics and CRM together is mostly known as CRM analytics or analytical CRM. Every customer-centric enterprise uses various CRM tools to analyze the customer profile. These CRM application uses Online Analytical Processing (OLAP)[8].

CRM lets an organization to modify business operations depending on the information obtained by analyzing the data. CRM makes it easier to choose the appropriate customers from a set of promising customers and Data Analytics helps the organization to propose the most tempting products to the current customers[5]. CRM analytics cater customer segmentation grouping (For example, grouping up customers based on most and least likely to repurchase a product), profitability analysis (an analysis which gives an idea of the most profit incurred over time with respect to individual customer), personalization/ customization (the action of modifying products for individual customers depending gon the information collected about them), event monitoring (for example, the event of reaching a particular dollar volume of purchases of a customer), what-if scenario (what are the chances of a customer buying a product which is similar to the previously bought product), and predictive modeling (for example, a comparative study of different product development plans as par their probable future success based on the customer profile). Collecting data and analyzing them are continuous are repetitive process and with time decisions are taken depending on the feedback from previous analysis and respective decisions[9].

Perks of CRM analytics are not only better and more beneficial customer relations with respect to sales and service but also enhancement in supply chain management (lower inventory and and speedier delivery) and therefore lower costs and more competitive pricing[4].

E. Why Apply CRM Analytics in Telecom Industry

CRM analytics in telecommunication industry is a rising but partially explored area. In today's competitive market, service providers always need to give something 'extra'- as in customized offers, own a predictive tool to identify the needs of the customer, make some offers which the customers cannot turn down, and so on. Telecommunication industry should not only be able to increase profit but also hold on to the existing customers. This industry contains a huge pile data but they are not being used properly. Lots of customer information in the form of network data, customer data, location, etc lies unused in the industry. In the coming years, telecommunication industry is going to be an important data source for innumerable other industries. There are more than one reasons for applying CRM analytics in telecommunication[8]:

1) ***Competitive Market:*** After years of unchanging market due to monopoly, now it is changing continuously rise in competition. Due to a large no of options in service providers, customers can switch any time. Hence telecommunication companies are using CRM analytics to gain reasonable dominance. By studying the statistical features and customers' actions, telecommunication industry can profitably modify their marketing strategies[10],[11].

2) ***High Churn Rates:*** Churn rate relates to the monthly or annual percentage rate of customers leaving the service provider. Due to a competitive market the churn rates are also high. At the beginning, when the telecommunication market was going through a rapid change, new customers were subscribing everyday and hence the churn rate was not a matter of concern. But with time as the market have matured, the churn rate have also increased. This high competition is making the telecommunication companies pay attention to their current customers and to find a way to make them stay. CRM analytics can be used to analyze the customer behavior to predict if a particular customer will leave and the reason behind it[10].

3) ***Massive Data Collection:*** The amount of data available in the telecommunication is massive, with the main product of the companies being the call. The thousands of calls created by the customer per day, per hour is one of the reasons for this huge data. Each call details record containing the time records, the details about the source and the destination, network details, and so on are stored in the database which is a very large one. Telecommunication enterprise also collect data about their customer and network data containing the status of the network components[12].

V. TYPES OF DATA IN TELECOMMUNICATION INDUSTRY

Before going into any technique in particular, the first step is to understand the data. Without understanding the data properly useful applications cannot be developed. In this paper we will be describing the three main types of data. Sometimes the raw data available are not directly usable in any

analytical tools, therefore it is necessary to transform the data to make it analyzable[13][14].

A. Call Details Data

It is a very detailed information of each and every call that are taking place. This detailed information is known as Call Detail Record (CDR). Considering the fact that over 300 million calls used to take place per day only for long distance, when the call details of several months are stored the number of records crosses over tens of billions.

All the important features of each and every call is included in the CDR. CDR are mainly real time data and are available for analysis almost immediately. These information are used for different purposes like billing data. The CDR is abridged into a single file portrays the customers' calling nature. Listed below are the aspects that are recorded in the call details of a customer based on whether it is the source or receiver:

- Average Call Duration
- % No-answer Calls
- % Calls from/ to a different area code
- % Weekday Calls (Monday- Friday)
- % of Daytime Calls (9 am- 5 pm)
- Average # Calls Received Per Day
- Average # Calls Originated Per Day
- # Unique Area Codes Called over certain Time period (P)

A customer profile consisting of these features consists of multiple promising applications. All of the attributes mentioned above, except the eighth one, are obtained from the basic data. But the exception of the eighth attribute is obtained by analyzing all the area codes over a certain period of time. To make the data mining process successful it is necessary to produce beneficial attributes. Though most of these attributes are constructed based on expert knowledge and common sense,it should contain some basic data analysis[15].

B. Network Data

The telecommunication network is composed of thousands and thousands of interconnected components which makes it a very complicated configuration. Each of these components or nodes are competent of producing a huge amount of data, that can be stored, in the form of error and status message which are used for network management. These messages contains at the least, a time-stamp, a string as a unique identifier for the generating hardware or software and a code containing the reason of the messages. Since the number of messages generated is too large to be handled manually, different analytical tools have been developed to examine the messages automatically and take necessary actions. The concept of time window is sometimes applied to achieve the goal of summarizing the real time data. These network data is mostly used for network fault isolation. Since each of the network component is able to do a limited self-diagnosis, they generate millions of error and alarm messages. These messages are needed to be analyzed in a timely manner to avoid any major network degradation[16].

C. Customer Data

Keeping a database of information becomes a necessity when the number of customers of telecommunication companies are in millions. The information contains of mainly the name and address and some other informations like service plan and contract information, credit score, family income and payment history. Since the customer data does not have any significant differences from other credit reporting agencies, hence those can be used to compliment the customer profile. These customer data are used for applications like fraud detection and market profiling.

VI. Data Mining Techniques Used in Data Analytics

Any choice of technique mainly depends on the type of parameters, need for analysis, data density to name a few. Similarly there is a wide range of techniques for predictive analysis which are used in the industry. In this paper we will be listing a few of those techniques and elaborate one of them (Clustering)[4].

A. Logistic Regression

It is a predictive analysis which is conducted when the dependent variable is binary. This technique is used to explain the data itself and the relationship of the dependent binary variable with one or more independent variable.

B. Linear Regression

It is the most basic type of regression which is used for summarizing and studying the relation between two continuous variables (one dependent and other independent).

C. Decision Tree

It is a graphical representation of every possible outcomes under any plausible conditions[2].

D. Clustering

It is a technique which divides a large group of data into smaller sub-group containing similar type of data. The main property of clustering is, it increases the intra-cluster similarity and decreases the inter-cluster similarity[2].

E. Neural Network

It is an approach that is loosely related to the biological brain and can be used to process information where various elements are interconnected in order to solve a particular problem[17].

F. Support Vector Machine

It is a supervised learning method which uses the concept of decision plane to separate set of objects having different class membership.

G. Others

Some of the other techniques that are used for the same purpose are Discriminant Analysis, Scorecards, Survival Analysis, Market Basket Analysis, Collaborative Filtering, Discrete Choice Modeling, Casual Models, and so on.

VII. DATA CLUSTERING

Clustering is an unsupervised learning technique which mainly includes bringing similar type of objects under the same class or group. It is a process which divides an available data-set into consistent groups depending on the available features and finds structure from a collection of unlabeled data[18].

Clustering Analysis is a rising research topic in data mining because of its range of applications, including image processing, computational biology, mobile communication, medicine and economics. The onset of many data clustering algorithms in the recent years and its use in wide variety of applications at a large-scale is the reason for the popularity of these algorithms. But the most prominent issue with the data clustering algorithms is its inability to be made regular. A developed algorithm might work on one type of data set but can fail to work on data set of other types. Many attempts has been made to standardizing the algorithms that will work on any type of data-sets but till date not much success has been achieved. There have been many clustering algorithms proposed so far but each of them have their own merits and demerits and not usable for all real situations.

According to the author, conventional pattern clustering comprises of the following steps[19]:
to the author, conventional pattern clustering comprises of the following steps[19]:

- Pattern Representation
- Definition of a pattern proximity measure
- Clustering or Grouping
- Data Abstraction (if required)
- Assessment of Output (if required)

Pattern representation comprises of different number of classes, various patterns that are available and the number, types and measure of the features available in clustering algorithm. Pattern proximity is determined by a distance function based on pairs of patterns.There are multiple ways to measure a distance among which the most simplest one is Euclidean distance that can be used to determine the dissimilarity of patterns. Then other similarity measures can be used to calculate the abstract similarity. The step of grouping can be performed in a multiple number of ways, like, hard or fuzzy output clustering, hierarchical clustering algorithms, partitional clustering algorithm, probabilistic and graph-theoretical clustering methods and so on. Data abstraction is the method of deriving a simple and compressed depiction of data-set. Cluster Validity Analysis is the evaluation of the cluster procedure's output.

There are different methods of clustering based on different criteria like density and interval, computing distance or a particular statistical distribution. Which method to be used or what criteria should the algorithm be based on depends on the data-sets and the requirements. The wide range of clustering algorithms can be broadly classified based on its implementation[18],[2]:

A. Partitioning Method

This method divides the data-set of n elements in k groups in such a way that every group contains at least one element, where each group represents a cluster and $k \leq n$. For example, K- means and K-medoids.

B. Hierarchical Method

This method is used when there is a need to group the data at different levels or in a hierarchical order by grouping into tree clusters. For example, Agglomerative, divisive, BIRCH.

C. Density-based Method

In order to overcome the problem of being unable to form clusters of arbitrary sizes, which is faced in algorithms based on distance, other algorithms have been developed based on the density.the concept behind this method is to let the cluster grow as long as the density exceeds the threshold. For example, DBSCAN, DENCLUE.

D. Grid-based Method

This method restrict the object space into a limited number of cells to form a grid. Due to the performing of all the operations on this structure the processing time is faster since it is only dependent on the number of cells in the structure. For example, STING, CLIQUE.

VIII. TABLE I
FEW CASE STUDIES USING DATA CLUSTERING

Title	Techniques Used	Assessment of the Techniques
Case Study on Cluster Analysis of The Telecom Customers Based on Consumers' Behavior[8]	Here the authors have used correlation analysis firstly to form a three dimensional model of the customer segmentation then applied Kohonen clustering method to form the clusters.	The user data of a regional telecom company in Liaoning has been used in this case study. The features of every cluster are used to mark strategy for the enterprise.
Customizing	Here the authors	Spectral Clustering is

Clustering algorithm for Data mining in CRM[1]	have applied spectral clustering technique on insurance data to overcome different customer related issues.	used on large network data-sets where the nodes are in the form of a graph. Hence it is much more useful than K-mean algorithm.
Using Data Mining for Mobile Communication Clustering and Characterization[20]	In this paper the authors uses K-mean algorithm on the customers' call records to determine their communication pattern to improve the business.	Here the K-mean algorithm has been used to investigate how it can characterize employers' communication patterns and how it effects the business activities.
Mining Profitability of Telecommunication Customers Using K-Means Clustering[6]	Here the authors have proposed a new K-mean clustering method with a RFM model which is used in telecommunication industry in Sri Lanka to evaluate the customer profitability.	In order to avoid the problem of initializing cluster centroids the authors have proposed a distortion curve to identify optimal initial centroids.

IX. CONCLUSION

The concept of CRM came into the picture in the early 1970's, which then got diversified into different sectors with the aim of attracting customers from different genres. With the amount of untouched data in telecommunication industry and the field being left unexplored to quite some extent it is like a gold mine for the data scientist. Much of the earlier works have been carried out using traditional techniques but in the recent years with the evolution in latest technology data science and data analytics have made their mark in the area of telecommunication industry. In this paper we have briefly discussed how Data Analytics and CRM is used in telecommunication industry. Due to the highly competitive market and the arising need to attract the customers with personalized services CRM analytics is a must. They mainly focuses on understanding the customer through communicating directly or analyzing their personal details or call details. CRM helps in better customer understanding, better understanding of the customer-facing operations, decision support, predictive modeling and bench-marking which in turn will help the telecommunication industry in their business prospective.

ACKNOWLEDGMENT

We would like to acknowledge KIIT University for the support in all respect w.r.t. the present work.

REFERENCES

[1] K. Rababah, H. Mohd, and H. Ibrahim, "Customer Relationship Management (CRM) Processes from Theory to Practice: The Pre-implementation Plan of CRM System", International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 1, No. 1, April 2011.

[2] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques, Morgan Kaufmann*, 3rd Edition, 2014.

[3] S. Mujawar, and A. Joshi, "Data Analytics Types, Tools and their Comparison", IJARCCE, Vol. 4, Issue 2, Page 488-491, February 2015.

[4] M. Kasthuri,V. Shanthi, and E. Manigandan, "Customizing Clustering algorithm for Data mining in CRM", IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555 Vol.6, No.2, Mar- April 2016.

[5] R. U. Maheshwari, S. S. Mahesan, Tamilarasan, and A. K. Subramani, "Role of Data Mining in CRM", International Journal of Engineering Research, Vol. 3, Issue 2, pp: 75-78, February 2014.

[6] G. K. Amoako, E. Arthur, C. Bandoh, and R. K. Katah, "The impact of effective customer relationship management (CRM) on repurchase: A case study of (GOLDEN TULIP) hotel (ACCRA-GHANA)", African Journal of Marketing Management, Vol. 4(1), pp. 17-29, January 2012.

[7] V. Haridasan, and S. Venkatesh, "CRM Implementation in Indian Telecom Industry- Evaluating the Effectiveness of Mobile Service Providers Using Data Envelopment Analysis", IJBRM, Vol. 2, Issue 3, Page 110- 127, 2011.

[8] D. Carmilovic, "Data Mining and CRM in Telecommunications", Serbian Journal of Management 3 (1) (2008) 61-72.

[9] B. Milovic, "Application of Customer Relationship Management Strategy (CRM) in Different Business Areas". Economics and Organization Vol. 9, N0 3, 2012, pp. 341-354, 2012.

[10] N. Kamalraj, and A. Malathi, "Apply Data Mining Techniques in Telecom Churn Prediction", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume- 3, Issue- 10, Page- 363-370, 2013.

[11] K. Dahiya, and S. Bhatia, "Customer Churn Analysis in Telecom Industry", IEEE, 978-1-4673-7231-2, 2015.

[12] S. K. Routray, S. Mishra, and L. Sahoo, "Adaptation of Fast Modified Frequent Pattern Growth approach for frequent item sets mining in Telecommunication Industry", American Journal of Engineering Research (AJER), Volume-4, Issue-12, pp- 126-133, 2015.

[13] Gary M. Weiss, "Data Mining in Telecommunications", Springer, Print ISBN 978-0-387-24435-8, 2005.

[14] M. Nadaf, and V. Kadam, "Data Mining in Telecommunication", International Journal on Advanced Computer Theory and Engineering (IJACTE), ISSN (Print) : 2319 – 2526, Volume-2, Issue-3, 2013.

[15] C. S. Hilas, P. A. Mastorocostas, and I. T. Rekanos, "Clustering of Telecommunications User Profiles for Fraud Detection and Security Enhancement in Large Corporate Networks: A case Study", Appl. Math. Inf. Sci. 9, No. 4, 1709-1718, 2015.

[16] G. Nagarajan, R. I. Minu, V. Vedanarayanan, S. D. Sundersingh Jebaseelan, and K. Vasanth, "CIMTEL-Mining Algorithm for Big Data in Telecommunication", G.Nagarajan et al. / International Journal of Engineering and Technology (IJET), Vol 7, No 5, Page 1709- 1715, Oct-Nov 2015.

[17] T. ZHANG, X. HUANG, J. TANG, X. LUO, "Case Study on Cluster Analysis of The Telecom Customers Based on Consumers' Behavior", IEEE, 978-1-61284-449-7/11, 2011.

[18] H. I. Arumawadu, R. M. K. T. Rathnayaka, and S. K. Illangarathne, "Mining Profitability of Telecommunication Customers Using K-Means Clustering", Journal of Data Analysis and Information Processing, 3, 63-71, August 2015.

[19] M. Chopra, "Analysis of Clustering Technique for CRM", I.J.E.M.S., Vol. 3(3) 2012.
[20] A. Bascacov, C. Cernazanu, and M. Marcu, "Using Data Mining for Mobile Communication Clustering and Characterization", IEEE, 978-1-4673-6400-3/13, 2013.