Telecom Churn Prediction Model Using Data Mining Techniques: Case Study in Pakistan

[1]Sufian Albadawi, [2]Khalid Latif,
[1]Muhammad Moazam Fraz
[1]School of Electrical Engineering and Computer Science, National University of Science and Technology, Islamabad, Pakistan

[2]Comsats Institute of Information Technology, Islamabad, Pakistan
email: sufian.badawi@seecs.edu.pk

***Abstract*—Recently several churn prediction models are being introduced that are able to predict possible churners to the service provider, in order to provide a retention solutions, i.e. offering them some compensations and special packages to keep them attached with their network. Predictive models can correctly identify possible churners. In this paper we are trying to implement a new prediction model that is uses data mining techniques on data obtained from a leading telecom company in Pakistan and then measure the generated benefits. The analysis shows a tremendous cost saving as compared to the expenses incurred in performing the churn management in a traditional way. We have employed the Decision Tree, Support Vector Machine and Neural Network classifiers on a six months subscriber's data sample to predict the churners' behavior.**

Keywords

***Index Terms*—Churner, Churn management Churn prediction, Neural networks, Support vector machine, decision trees, customer satisfaction, profitability, confusion matrix., classification, prediction, data mining, clustering**

1. INTRODUCTION

Not only the Telecom. Industry, is working on identifying churners as many other industrial researchers are working in it as a hot topic in the management science. Any proposed Churn prediction model has to predict potential churners accurately so that energies used to retain these customers are not wasted on misclassified users, who were not even churners. So churn prediction model has to be accurate. In this paper we have merged and implemented some of the most famous methods used for churn predictions and then compered them to find one winner so that people can focus on this model in order to get better results. The main focus in this paper was on telecom industry churners.

Prediction based Churn Management is a being practiced globally by many telecom companies and other businesses. Continuous efforts and resources are being invested in the research and development activities to obtain the maximized level of prediction accuracy. Business and technology researchers and professionals are trying to evaluate several possible churn prediction models according to their business and research targets. During the last decade enhanced the importance of prediction for churn reduction and customer retention [1].

The data mining techniques used for prediction produce results from the appropriate database variable sets. Data mining provides useful clustering and classifications of customers' data and the prediction results may also be categorized based on demographic and behavioral factors. Several data mining techniques are in use for the classification/clustering of data to predict future churners. Among available techniques, Decision Tree (DT), Support Vector Machine (SVM), Neural Networks (NN), Genetic Algorithms (GA) and Fuzzy Logic (FL) are more common for churn prediction.

This paper is divided in 6 major sections i.e. Introduction of the focus of study, Churn Introduction which also includes PTCL Churn Process, measure and analysis, the 'New Proposed Model' section is to elaborate the results of data analysis which is the bases for this case study, Future Work and Conclusion.

2. LITERATURE REVIEW

The business operators in telecommunication industry depend on their customer base and market share. It becomes a primary for the business managers to stabilize the profitability by inviting new customers while retaining the existing. Customer retention and churn is becoming a critical factor for sustainable presence of a telecom operator. To define strategies for retaining customers and reducing churn, the decision makers depend on the predictions based on a large amount of customer's data. The most important factor before designing and implementing a prediction model is carefully determining the data usage and how it should be used. It will determine the group of customers to offer suitable incentives for accomplishment of best churn reduction targets. The level of complication is high while identifying the useful data and finding the desired outcomes. To find the optimistic results, it will require understanding the concepts of fundamental data science.

The churn and customer retention is a major subject for using data mining in telecom business. Businesses like telecom and financial institutions were the primary users of data mining for the prediction of churn. Data mining technologies emerged as effective tools for the Identification of the churn indicators in the data. The goal for such predictions should be to increase the weight of meaningful and useful knowledge [1].

As this paper is focusing on the prediction of churners, so the reasons for churn are not in the scope. The study is more specifically about the prediction of customers who are not yet churners but could be in the future. This prediction is useful for targeting such customers to reduce the churn in future.

Instant churn prediction, enables better business decision making that follows an approach to identify and retain the risky customers who are about to churn using different churn prediction methods like Generalized Additive Models (GAM) [2] . trying to retain telecom subscribers with special offers designed for those customers who have been classified as potential customers to churn to another telecom operator with high likelihood of doing so, for that a model called data mining by evolutionary learning (DMEL) is introduced by Wai-Ho that merges machine learning with genetic algorithm that enables recursive rule creation with enhanced interesting measures' that more efficiently identify and classify the churn patterns [3], following the subscribers that are about to leave the service by satisfying their needs under resource constraints is churn prediction, and aims to detect customers intended to leave a service provider. A study of churn prediction via SVM has been implemented in bank industry and helped in improving the prediction model that enable reduce churn while trying to increase customer base has reduced loss and increased the created value [4] this study discusses commercial bank customer churn prediction based on SVM model, and uses random sampling method to improve SVM model, considering the imbalance characteristics of customer data sets. The results show that this method can effectively enhance the prediction accuracy of the selected model

Churn management is a hot area of research in management other research disciplines in the recent few years, tried to identify optimal models for churn prediction and highlighted ways of reducing the churn in the organizations.

There are many DM techniques that can be used in classification and clustering customer data to predict churners in the near future [1] such as neural networks, decision trees, and logistic regression [5].

From surveying what different scholar wrote about churn management in telecom, we can summarize the churn management as if we are analyzing it following the five steps of customer behavior analysis model [6].

So we target specifying the attributes that can be used uniquely for the identification of the churned customer and tracing his customer behavior during his life period till he become a churner, and we should be able to detect and study the phenomena of how many times he has churn and come back to the service and what are the factors that are effecting that.

To do so, we need to try to find as much attributes in telecom data and make a dataset, After identifying the dataset, We take the data from different sources for an identified period of time where we have decided to study the customers churn behavior and then by having this dataset churned customers we can go to next step for further analysis to finally build a model which can predict the churn rate of different types of customers and identify churn patterns and what are the best actions that we can do to retain this segment of customers.

Essam Shaaban et al in [10] stated that there are two types of churn presented in telecom industry, voluntary and involuntary. Involuntary churners are not important as they are decided by telecom, the one's they don't want as their customers. Voluntary churners are the one in focus here.

In voluntary churners, Incidental churn are not preplanned. Deliberate churn are the main churn we are interested in, they can leave the network because of poor service, financial issues, social or psychological factors [1].

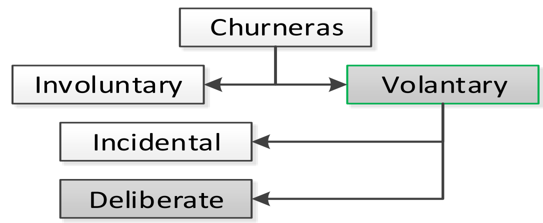

Fig. 1. Churner Types [1]

Michael Mozer et al in [5] stated that Churn is influenced by several factors that can be summarized in table 1:

Table 1: FACTORS INFLUENCING SUBSCRIBER SATISFACTION

Influencing Factor	**Importance**
Call quality	21%
pricing and charging Rates	18%
Corporate Capability	17%
Customer complaints	17%
Customer communications	10%
Telecom. Devices/ Handset	4%
Roaming / Coverage	7%
Cost of roaming	3%
Billing	3%

In the market and previous literature several data mining models are used and presented. And they are classified to two type of predictive.

Firstly are using traditional techniques: Decision trees and Regression analysis. In decision tree first we build the tree and keep partitioning until all partitions contain the same data. Then we prune the tree removing branches with the highest error, this part enhances the tree accuracy. Regression analysis uses past data to analyze what are the conditions when a customer leaves a network and the reasons behind those conditions.

Other methods being used is soft computing techniques: Neural networks, Fuzzy logic. Neural network also have various ways to implement within themselves and they predict churns by using likelihood. Fuzzy logics are very easy to implement but their results are not very good and they are usually not used to find churners in telecom industry [1].

A. PTCL Current Churn prediction Process

As the leading telecom operator in Pakistan PTCL needs to maintain and grow its market share. Like other telecom companies, PTCL is also concerned about churn due to the technology and market competitions. PTCL has defined a Credit Monitoring Policy (CMP) for overdue accounts and defaulted customers. The use of CMP is to setup a tolerance level of bill non-payments against the offered services.

- *Credit Monitoring Policy Steps*

One-way suspension (TOS) of service if a customer doesn't pay the one a month's bill by due date and the amount is less than credit limit.

(i) *- Recovery follow-up after 15 days of TOS, the service remains one-way suspended.*
(ii) *- Two-way Suspension (POS) after one after TOS, customer is intimated before POS.*
(iii) *- Cessation and archival of all services, transfer to defaulter ledger.*

The PTCL's churn prediction is based on the service usage statistics of customers and CMP. The data generated by the exchanges provides details of monthly incoming and outgoing calls of each customer. Customers with no outgoing calls are considered for churn analysis.

B. Traditional PTCL Churn prediction method Process Steps

1. Customer Care department will develop a listing of zero usage by working with IT and Commercial department.
2. The list will be refined as:
   - Segregation of Landline and Broadband customers
   - Filter out numbers having incoming calls
   - Filter out numbers having 'Incoming only facility' only
   - Filter out number having 'Safe Custody' state
   - Filter out numbers in 'Pending for closure' state
   - Identify CRM complaints history for a refined list

3. List of numbers with incoming calls to be shared with Outbound Contact Center (OBCC)
   - OBCC will dial out these customers and capture the response (Line faulty, Multiple lines, applied for closure, safe custody, using for incoming only etc)
   - The response report to be shared with the relevant headquarter and zonal management.
4. SMS/IVR/Email auto messages for reasons of no usage sent to customers having incoming calls and taking the feedback through SMS/IVR/Email.
   - IT captures the feedback received and the reports to be shared with stakeholders.
   - List of customers who have not registered feedback shared with OBCC for contacting on customers' primary and secondary numbers. Reports to be shared with the stakeholders.
   - List of customers who could not be contacted in previous steps is shared with the regional teams for contacting customers. The reports will be shared with stakeholders.
5. Random sampling technique is used to verify the complaints resolution claims/feedback of customer by the Quality Assurance department.
6. A weekly progress summary on Zero Usage Activation is generated at a regional level which is consolidated as a summarized result.

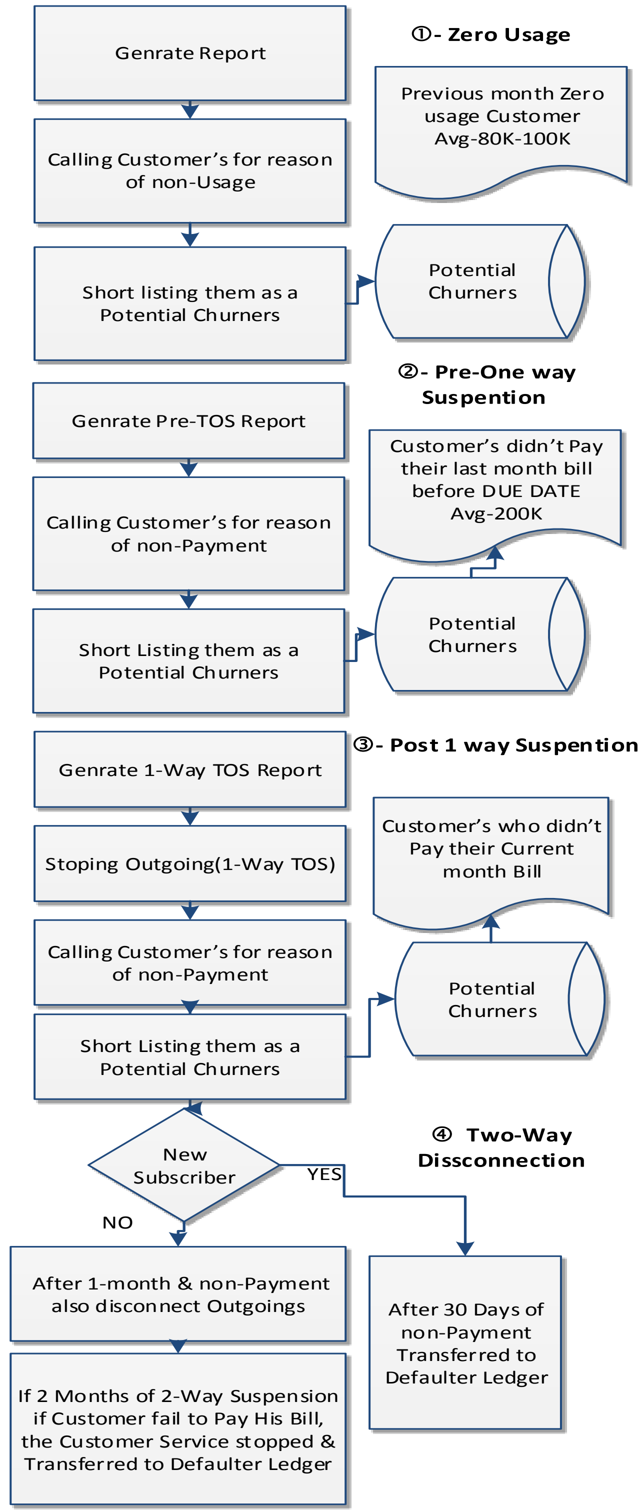

Fig. 2. PTCL Churn Prediction

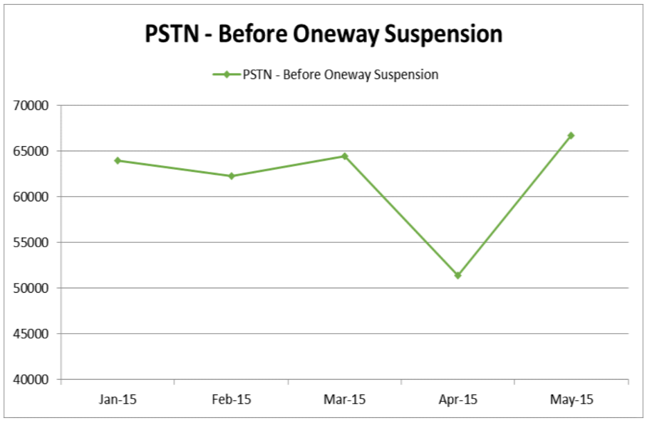

Fig. 3. Churn figures before one way Suspension

3. PROPOSED METHODOLOGY

A model that works in 9 steps is proposed, those steps are:

1- Identifying the churn Problem and its needed data inputs.
2- Formalize the dataset structure and the ETL processes for having its pre-mining steps.
3- Identifying the moving window for frequently capturing the dataset.
4- Formalizing the process of mining steps and confirming the mining methods.
5- Running the classification, generating, and filtering the churners.
6- Formalizing the clustering method and the alternative retention methods for the cluster.
7- Generating the clusters and validating the results.
8- Revising the churn patterns with fewer frequencies and recommending a course of action against frequent customer churn behavior.
9- Reviewing the discovered knowledge.

Matlab and WEKA tools are used to analyze the case study. The Total 8100 Tuples of data were divided into two datasets, 5400 Tuples for training and 2700 for testing. Out of 5400 training instances 5130 were labeled as non-churners and 270 as churners. Total of the attributes count was 23 attributes for this dataset. The same dataset is used to predict the output by using three different models: Decision Tree, Neural Network and Support Vector Machine The source of this dataset is extracted from the live systems of billing, customer care, complaints management, and interconnect billing during a period from Nov-2014 till Mar-2015 for customers from one telecom region in Pakistan.

3.1. Data Preprocessing

Normally to work on churn data we need to cover a wide range of the customer behavior attributes that varies between Customer Service and complaints including the nature of complaints and resolution, Billing details like base fee, roaming charges, charge for minutes beyond prepaid limit, Market details like competitor rate plans, Application for Service like credit

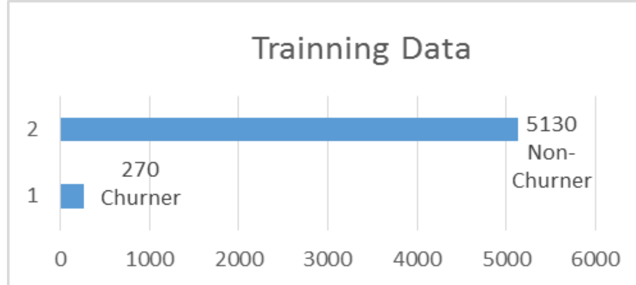

Fig 4. Training Data Set

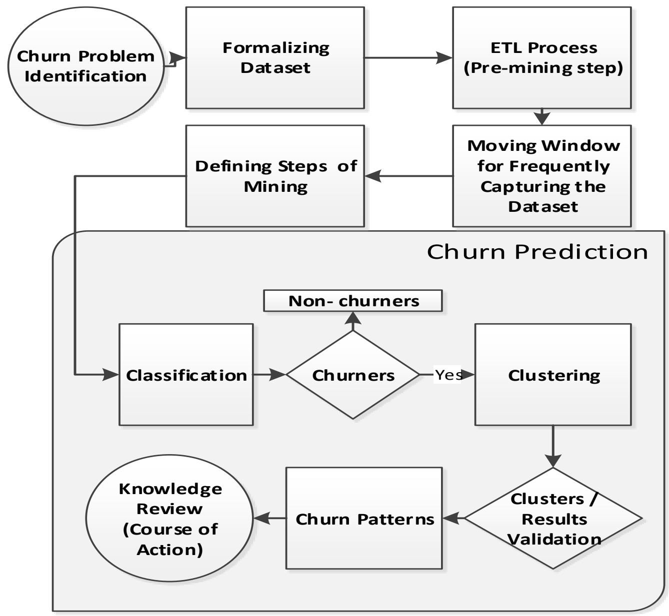

Fig 5: The Proposed Churn Prediction Model

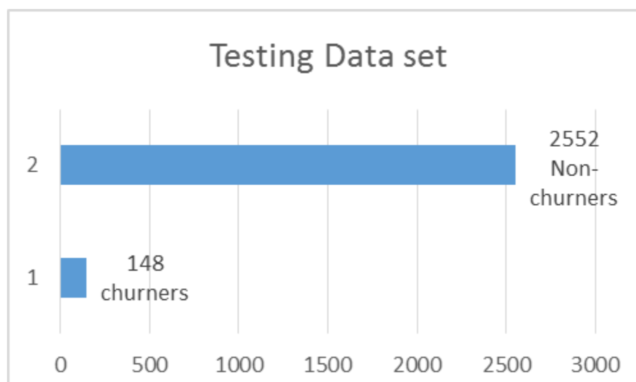

Fig. 4. Testing Data Set

history, customer classification (Consumer, corporate) , rate plan, active services (number, type, cancellation dates, avenue of activation), telecom device type, Demographics like average income, population density, Age, Marital status, Network like quality of service, usage patterns (number of calls, duration of calls, incoming calls, outgoing calls, location of calls, peak/off peak), and dropped calls [7].

Before applying any data mining techniques, we need to have the dataset ready to be mined for useful information. We have followed the following preprocessing steps:

We have analyzed the needed data and found that we need to refer to the systems (billing, receivable, interconnect billing, CRM, reports, customer care records of previous churn analysis work).

Six months data is collected to cover a period that contains churn behavior (as churner needs in average 3 months to be transformed from non-churner to churner). Obviously as this interval in Telecom Company is huge we have reduced the scope to the data of a few telecom exchanges to enable us to process this data for the data set preparation. Several challenges meet like getting the table space full to gather the data from all the systems, interactions with different system teams to get the data as needed in the format requested, till finally we gathered and collected the data.

As we know that the gathered data for analysis is a raw datasets that has to be understood and analyzed carefully. With heavy efforts of preprocessing steps we still got a high error rate (presently in this paper varies around 21% to 28%). That is due to several reasons including that dataset preparation are happening the first time for this purpose, labels are taken from decisions based on previous traditional ways of churn prediction. Following preprocessing steps that can be applied but not mentioned in this paper.

3.1.1. Missing and noisy Data

Firstly following careful analysis the noisy and missing data needs to be smoothed, outliers in a dataset should be excluded. As a model trained on such noisy data could face over fitting problems during the prediction of unseen data, we have noticed that a main customer attributes are not available in customer records that needs to be gathered initially from all customers like the customer date of birth , marital status and customer gender.

3.1.2. Data Integration

Secondly integrating data from all the above mentioned systems was a real challenge that we have handled to merge the data in one homogeneous dataset, and maintaining the needed keys to link data from different systems for the same billing month periods.

3.1.3. Dimension Reduction

Thirdly we have worked on reducing the dimensions to have one data set per customer per average month for the customer's behavior throughout the customer life cycle from availing the service till data mining date.

3.1.4. Data Normalization

Data was not normalized before making predictions. Data normalization step can normalize the data well; hence can play a good role in achieving good results.

4. CONTRIBUTION

The main idea in this paper is to demonstrate a practical case study that is showing the benefits of shifting churn prediction and reporting from the old traditional ways in telecom industry to data mining methods. Basically we have tried to make Telco industry aware of the best model available in the market to predict churners, so that the telecom industry management don't waste their time and money on wrong customers, who were not even churners.

4.1. Results and Analysis

Pakistani Telecom company data is collected for a period of six months and passed to the proposed method that classifies the customers using three data mining classification models, Neural networks, SVM, and Decision trees with comparison of the classification results.

The confusion matrix of the generated models in Table 2 below shows the results acquired after this case study on a specific data.

Table 2: DECISION TREE CLASSIFICATION CONFUTATION MATRIX

	Actual	Prediction	
		Churner	Non-Churner
Decision Trees	Churner	210	325
	Non-Churner	430	1735

Table 3: NEURAL NETWORKS CLASSIFICATION CONFUTATION MATRIX

	Actual	Prediction	
		Churner	Non-Churner
Neural Networks	Churner	360	355
	Non-Churner	288	1697

Table 4: SUPPORT VECTOR MACHINE CLASSIFICATION CONFUSION MATRIX

	Actual	Prediction	
		Churner	Non-Churner
SVM	Churner	410	298
	Non-Churner	265	1727

Error rate and accuracy for each model are shown in the table5

Table 5 : ACCURACY AND ERROR RATES

Method	Accuracy	Error Rate
Decision Trees	1945/2700 72%	755/2700 28%
Neural Networks	2057/2700 76%	643/2700 24%
SVM	2137/2700 79%	563/2700 21%

By Looking at table 5, we can conclude that Support Vector Machine as well as NN are the best performing algorithms out of these three as they estimate more churners correctly than other methods.

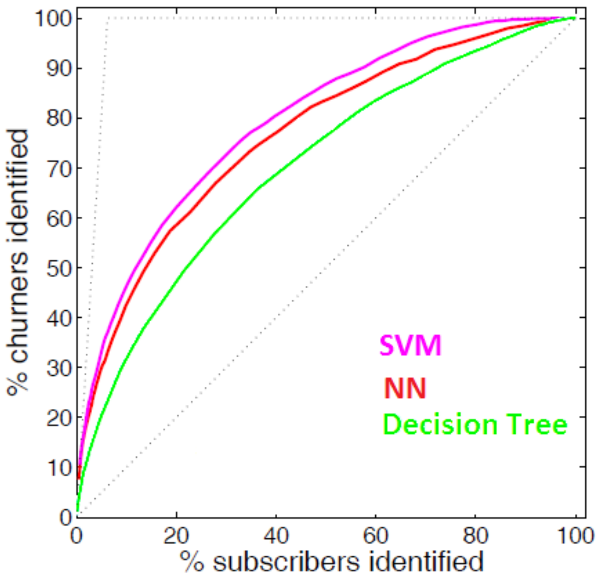

Fig 5. Lift Curve of the Three Models

Figure 5 indicates that SVM model is more discriminative than NN or Decision tree models as it is the closest to top left corner.

5. CRITICAL ANALYSIS

There is no doubt that the current Churn prediction activities are working towards fulfilling the business needs in the organization. Yet it is clear that referring to data mining techniques and methods to identify the possible churners is more precise, unbiased and optimal in churn management.

By applying Clustering of the gathered dataset in WEKA using K-means model, table-6 shows for the resulting clusters generated

Table 6 : Kmeans (K=2) Clustered Instances

Cluster	Tuples count	%
0	5500	68%
1	2600	32%

That indicates that the dataset gathered has an indication of 32% of the potential churners that needs deeper work and review for the customer offerings and customer retention efforts.

5.1. Churn Prediction Criteria

In churn prediction analysis we should state when we consider the record of a customer to be a churner record i.e. what are the criteria for the customer to be churner? The criteria includes a customer who used to use the service normally and pay his bills regularly then he decided voluntarily to revoke his service regardless of whether he did inform about this intention directly or not, his usage details will be indicating this intention like stopping initiating a calls or receiving a calls maximum for 3 months.

5.2. Influencing Factors

By running classification on the data using SVM regression in WEKA we clearly analyze the factors influencing churn out of the studied 23 attributes.

== Classifier model (full training set) ===

weights (not support vectors):
- 0 * (normalized) GENDER
+ 0 * (normalized) MAR_ST
- 0.0008 * (normalized) M_IN_SM_MOU
+ 0.0003 * (normalized) M_OUT_SM_MOU
- 0.0003 * (normalized) M_SM_MOU
- 0.0005 * (normalized) M_IN_OTH_MOU
+ 0.0002 * (normalized) M_OUT_OTH_MOU
- 0.0003 * (normalized) M_OTH_MOU
- 0.0011 * (normalized) M_IN_MOU
+ 0.0004 * (normalized) M_OUT_MOU
- 0.0005 * (normalized) MOU
- 0.0002 * (normalized) Change_MOU
+ 0 * (normalized) M_SMS
- 0.0028 * (normalized) M_M_REV
+ 0.0009 * (normalized) ASS_PROD
+ 0 * (normalized) ASS_SER
+ 0 * (normalized) M_CC_CALLS
+ 0 * (normalized) M_DROP_CALLS
- 0.0014 * (normalized) COMPLAINTS
+ 0 * (normalized) NO_CHNG_TIF_PLAN
- 0.0007 * (normalized) PHONE_NUMBER
- 0.0002 * (normalized) MAX_CALL_DISTANCE
+ 0.0005

The above WEKA results shows that the mean revenue is the most influencing factor on the classification compared to the rest of parameters then customer's complaints, followed by Mean outgoing calls Minutes of Usage.

We recommend further analysis and studies on those factors during targeting customer's retention as well as reviewing the gathered dataset attributes and adding further influencing parameters from table 1 that will optimize the data set selection.

6. FUTURE WORK

6.1. Model for assigning retention strategies

Presently no such type of model exists that can assign suitable retention strategies for each churner type. In this research paper, a classification and clustering problem is solved but didn't provided any model that can suggest any suitable retention strategies as per churn cluster. That model is expected to try to provide new churn prediction retention model that will use the predicted and clustered data to assign a suitable retention strategies for each churner type.

6.2. Bagging and Boosting Approach

The method of bagging and boosting is a very good tool that helps in optimizing the suggested model.

Especially once outliers are detected and removed, a more promising result can be achieved from bagging and boosting techniques with a higher accuracy results in the generated model.

7. CONCLUSION

In this paper, we have presented the optimal use of data science in churn prediction and management by using a case study. More specifically, the paper presents the techniques to reduce the churn and increase the profitability and consequently increasing the customer satisfaction through the data mining techniques instead of traditional methods of churn management.

Although we have presented the method for optimizing the churn prediction, and majorly focused on the comparison of the classification methods on the collected dataset, and influencing factors, extra mining results would be achieved by proceeding in the cause-effect analysis of identifying the most influencing attributes from customer details elaborated in table 1, and finally clustering and pattern finding out of the customer churn logged behavior. Soon in the future, the technology will be unleashing a revolution in management sciences based on the practical algorithms that is matured by artificial intelligence data mining, and machine learning techniques.

8. References

[1] B. H. a. Y. S. a. Q. W. a. X. Zhao, "Prediction of Customer Attrition of Commercial Banks based on \{SVM\} Model," *Procedia Computer Science,* vol. 31, pp. 423-430, 2014.

[2] r. U. o. H. r. Rami Mustafa A Mohammad, "Machine Learning Reposatory," Center for Machine Learning and Intelligent Systems, 2015. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Phishing+Websites. [Accessed 1 May 2105].

[3] K. C. a. D. V. d. Poel, "Integrating the voice of customers through call center emails into a decision support system for churn prediction," *Information \& Management,* vol. 45, pp. 164 - 174, 2008.

[4] W. L. W. Z. A. A. A. Nengbao Liu, "Handling Class Imbalance in Customer Behavior Prediction," *Institute Center for Smart and Sustainable Systems (iSmart) ,* 2010.

[5] C. M. Michael, W. Richard and D. Robert, "Predicting Subscriber Dissatisfactionand Improving Retention in theWireless Telecommunications Industry," *Athene SoftwareBoulder.*

[6] S. V. Kiranjot Kaur, "Enhanced Boosted Trees Technique for Customer Churn Prediction Model," *Vol. 04, Issue 03 (March. 2014), ||V5|| PP 41-45,* vol. Vol. 04, no. Issue 03 (March. 2014), pp. ‖V5‖ PP 41-45, Mar 2014.

[7] S. Joshi, "Customer Experience Management: An Exploratory Study on the Parameters Affecting Customer Experience for Cellular Mobile Services of a Telecom Company," *Procedia - Social and Behavioral Sciences,* pp. 392 - 399, 2014.

[8] F. P. &. T. Fawcett, Data science for business, USA: O'Reilly, 2013.

[9] Z. N. S. a. Z. P. Faizurrahman, "Managing Customer Retention of Hotel Industry in Malaysia," *Procedia - Social and Behavioral Sciences,* vol. 130, pp. 379 - 389, 2014.

[10] Y. H. A. K. M. N. Essam Shaaban, "A Proposed Churn Prediction Model," *International Journal of Engineering,* vol. Vol. 2, no. Issue 4, pp. pp.693-697, June-July 2012.

[11] Y. H. A. K. M. N. Essam Shaaban, "A Proposed Churn Prediction Model," *International Journal of Engineering,* vol. Vol. 2, no. Issue 4, pp. pp.693-697, June-July 2012 2012.

[12] K. a. B. D. a. P. D. d. Coussement, Preventing Customers from Running Away! Exploring Generalized Additive Models for Customer Churn Prediction, Springer International Publishing, 2015.

[13] W.-H. a. C. K. C. a. Y. X. Au, "A Novel Evolutionary Data Mining Algorithm with Applications to Churn Prediction," *Trans. Evol. Comp,* vol. 7, pp. 532-545, 2003.

[14] A. Ã. a. M. T. A. a. M. Argan, "The Effect of Mobile Service Quality Dimensions on Customer Satisfaction," *Procedia - Social and Behavioral Sciences,* vol. 99, pp. 428 - 438, 2013.

[15] L. Alberts, Churn Prediction in the Mobile An application of Survival Analysis in Data Mining, Maastricht University, 2006.

[16] 4th International Conference on Marketing and Retailing 2013, "A Business Intelligence Perspective for Churn Management," *Procedia - Social and Behavioral Sciences ,* vol. 109, pp. 51 - 56, 2014.