Churn Prediction Retention Framework

Essam Shaaban

Aboul Ella Hassanien

Abstract- **Churn prediction is considered a big issue in the Telecom market because customer acquisition costs five to eight times than retaining an existing one. Customer retention is the action that a service provider undertakes in order to reduce customer dissatisfaction and decrease the probability of leaving a service provider. This paper proposes a churn prediction retention framework based on data mining techniques. Deploying the proposed framework into a business intelligence system can help in enhancing the efficiency of customer relationship management. Moreover it can help customer churn management department to easily predict and retain the expected future churners in many business areas. The proposed framework mainly composed of seven phases and sub-phases. A case study is demonstrated to evaluate the framework throughout 5000 customers' records in an anonymous telecom company.**

Keywords- **Churn prediction, customer retention, data mining, clustering, classification**

I. Introduction

The purpose of prediction is to anticipate the value that a random variable will assume in the future or to estimate the likelihood of future events [1]. Churn prediction is a useful tool to predict customer at churn risk. Customer Churn Management (CCM) becomes one of the most critical success factors mainly due to higher acquisition costs for new customers [2]. The term churn management has been adopted to define customer turnover [3]. Besides; CCM is the concept of identifying those customers who are intending to move their custom to a competing service provider. Retaining an existing customer from being a churner is the best core strategy to survive in industry as well as it is becoming a common knowledge in business [4], [3]. The prevention of customer churn through customer retention is a core issue of Customer Relationship Management (CRM). Customer retention is defined as the practice of working to satisfy customers with the intention of developing long term relationships with them [5]. Retention is a result of customer desertion. Urquizo [6] stated that the ways in which customers leave are; credit does not sufficiently help customers achieve their goals,

Essam Shaaban
Faculty of computers and information, Beni-Suef University
Egypt

Aboul Ella Hassanien
Faculty of computers and information, Cairo University
Egypt

customers are being negatively affected and are having difficulties making payments and customers find alternative service providers. Regarding this research, customer retention can be defined as the organizational plans, actions and activities to retain potential customers by developing, maintaining and maximizing mutually beneficial long-term relationships [7]. It is supported by many factors, including: high customer churn rate, evolutionary needs and behavior of customers, increasing national and international competition, mature markets, and inappropriateness of traditional retention analysis tools to tackle the problem of having a churn prediction retention model in Business Intelligence (BI) applications that can efficiently identify churners in telcos and provide a retention solution. The aims of this paper are to analyze the existing prediction and retention models to highlight their capabilities and limitations. Then develop a framework for Pre-paid mobile churn prediction and retention.

This paper is organized as follows. Section II describes the theoretical background in charge of prediction models, customer retention, and limitations of churn prediction and retention models. Section III shows in a detailed format the proposed churn prediction retention framework. Section IV provides the results of an implemented case study. Finally; conclusion and future work are presented.

II. Theoretical Background
A. *Prediction Models*

A predictive model is defined as one that takes patterns that have been discovered in the database, and predicts the future [8]. Predictive modeling is mainly concerned with predicting how the customer will behave in the future by analyzing their past behavior. As Figure 1 depicts, prediction models can be classified into two types; traditional models and soft computing models.
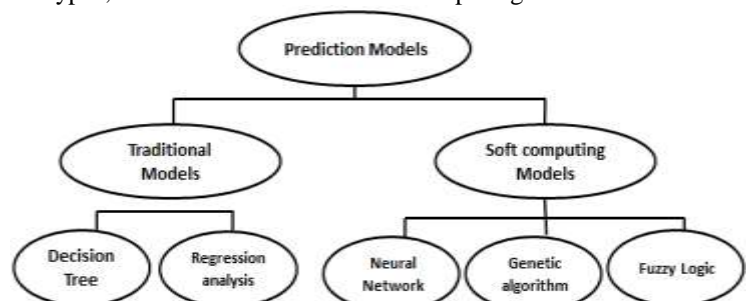

Figure 1: Prediction models

The most popular traditional prediction models are; Decision Tree and Regression Analysis. Decision tree is the most popular type of predictive model. It has

become an important knowledge structure, used for the classification of future events [9]. Regression Analysis is based on supervised learning models. Regression models deal with a dataset consisting of past observations, for which both the value of the explanatory attributes and the value of the continuous numerical target variable are known [1]. On the other hand Hadden et al. [3] stated that soft computing is a consortium of methodologies such as Fuzzy Logic (FL), Neural Network (NN), and Genetic Algorithms (GA) that work cooperatively and provides, flexible information processing capabilities for handling real-life problems. Also Pal and Ghosh [10] argued that exploiting the tolerance for imprecision, uncertainty, approximate reasoning and partial truth in order to achieve tractability, robustness, low solution cost, and close resemblance with human-like decision making is the aim of soft computing technologies.

B. Customer Retention

Customer retention is defined as the practice of working to satisfy customers with the intention of developing long term relationships with them [5]. Table 1 summarizes an investigation about the existing retention strategies in the literature.

Table1: retention strategies

Year	Author	Area of interest
2013	Palto Ranjan Datta [11]	Retail
2012	Zhi, Xia and Gan [12]	Telecom
2011	Lee et al. [2]	Marketing programs
2011	Li, Huang, Li and Zhang [13]	Telecom
2011	Roberts [14]	Insurance
2005	Xevelonakis [15]	Telecom
2005	Angela Haran [16]	Insurance

Developing such a model because of some important reasons such as; acquiring new customers is more difficult than retaining the existing ones, the increased profit that customer retention strategies create minimize the cost of selling new products or services because it is easier to sell to existing customers than selling to new ones [17]. Moreover; new customers are more difficult to find, and they often pay less than existing customers do. Besides; the obvious direct benefits that can be realized by adopting customer retention strategies, there are many indirect benefits such as attaining valuable customer feedback that can be used to improve business operations and Word-of-mouth (WOM).

C.   Limitations of churn prediction and retention models

The variations among the previous studies of churn prediction provide a great fact which is; results of prediction model are based on the dataset in hand. In other words, the difference in results between the most

commonly used prediction models is based on the number and nature of attributes. Regarding the existing retention literature, one can argue that there are obvious differences between the markets across countries, so the results and the outcomes of such researches cannot be generalized. Verbeke et al. [18] asserted that much of the customer retention studies focused on predicting churn and paid less attention to the causes of churn. In addition, they indicated that the lack of comprehensibility is one of the major issues in customer retention models, and confirmed the need to build models that are more comprehensible. Many on the existing customer retention studies [19], [20], [21], and [22] are focused on investigating the loyalty of customers throughout questionnaires to measure the customer satisfaction. So there is lack of studies regarding usage of computer techniques to find the best retention solutions. Also most of the existing models discuss only churn prediction with neglecting retention solutions. Hassouna [7] and Ahn et al. [23] pointed out that although the large number of studies done on customer retention, few of these studies construct a comprehensive view of the relationship between the different factors and customer retention. So the proposed churn prediction retention framework in this paper makes relationship between the predicted churners and their behavior with the provided retention solution

III. The proposed Churn Prediction Retention Framework (CPRF)

The proposed framework is named Churn Prediction Retention Framework (CPRF). It **is** based on some of the existing churn prediction models, knowledge discovery models, and retention strategies. These models and strategies have been described previously in chapter two. As figure 2 depicts, the CPRF framework is generally composed of seven phases which are; identify business domain, Data preprocessing, investigate data set, prediction, testing, retention, and deployment. Moreover; the prediction phase is consist of two sub-phases which are classification and clustering. Besides; the retention phase is constituted to classify churn attributes, available solutions or solution bundle and assign solution. The following section describes in details the phases and sub-phases of the proposed framework.

1. Identify Business Domain Phase. This phase is used to understand and identify clear business objectives, goals, and requirements.

2. Data Selection Phase. This includes data collection from the available data sources whether it is a DW, databases, spread sheets, or any other data sources. Moreover; this phase includes preparing the data set for the churning prediction retention process. It also includes data and attributes selection, cleaning, construction of new attributes, and data transformations.
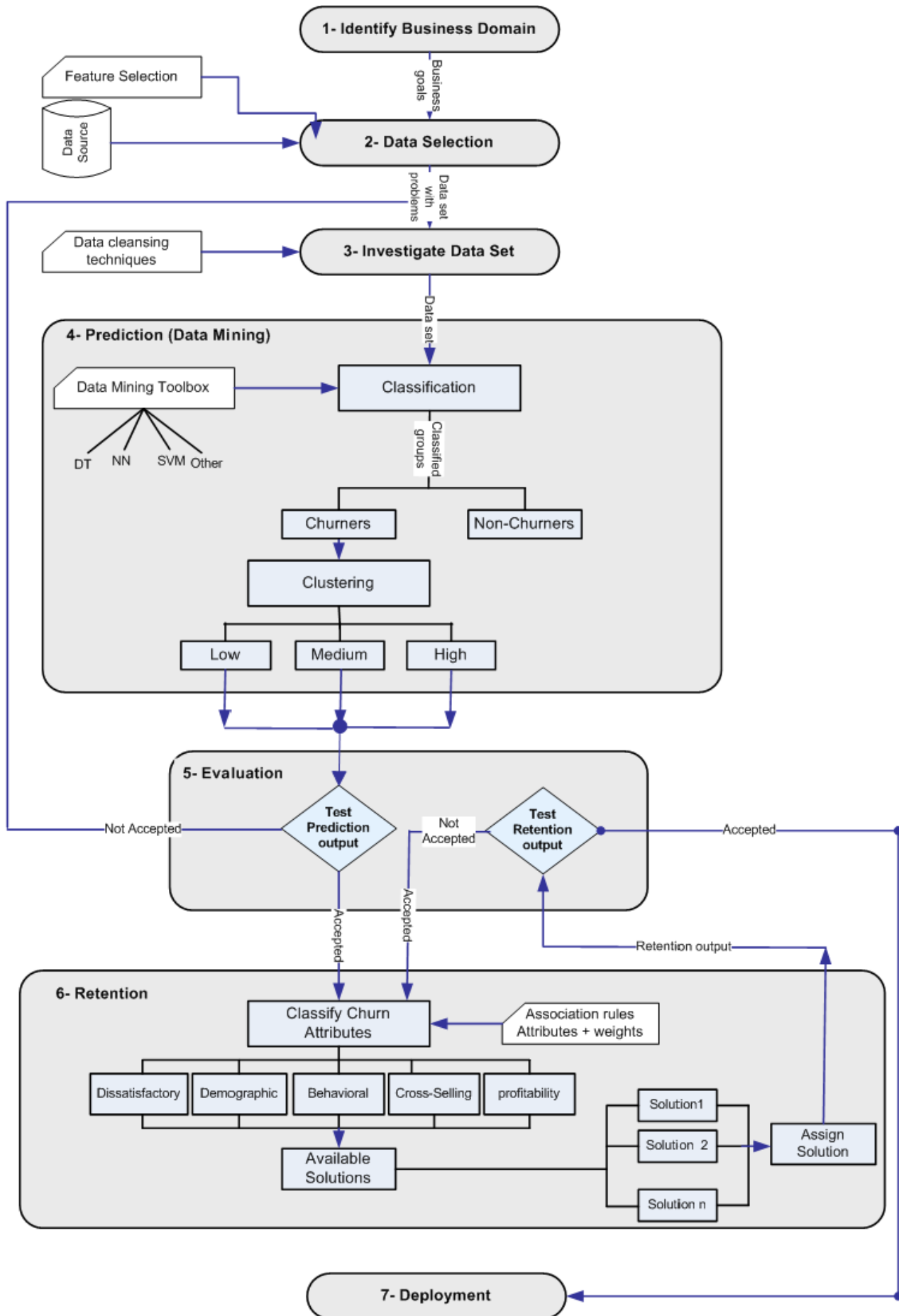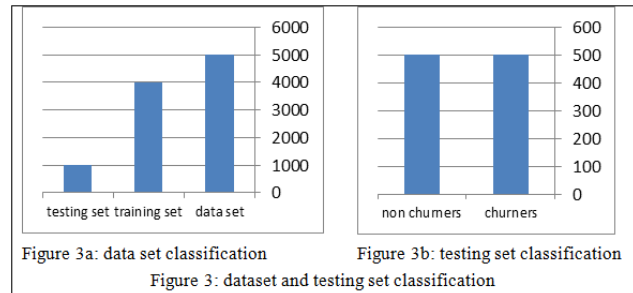
Figure 2: Churn prediction retention framework   13

3. Investigate the Data Set Phase. This phase concentrates on the data quality issues such as accuracy, completeness, uniqueness, timeliness, and consistency.

4. Prediction (DM) Phase. The main objective of this phase is to predict and produce the expected churners. This phase is composed of a set of sub-phases which are classification and clustering. The inputs for this sub-phase are: the final data set after solving the data quality problems and issues, and the algorithms, methods or technique used for accomplishing DM process. The prediction sub-phases are:

  i. *Classification*. Classification is a supervised way of learning. The inputs for this sub-phase are the data set and the classification technique. The used DM technique can be accessed from the DM toolbox. The toolbox means that there many types of methods that vary from traditional methods and soft computing methods.

  ii. *Clustering*. Clustering is a division of data into groups of similar objects. The input for this sub-phase is the churners that were produced from the classification sub-phase. The outputs of this sub-phase are three different clusters based on the specific threshold, criteria or attributes used for the clustering process. These three different clusters are low, medium, and high. Each of the produced clusters is tested in the next testing phase in order to be assigned a retention strategy in phase 6

5. Evaluation phase. Evaluation is a means for gaining better understanding of what we do and the effect of our actions in the work environment. This Phase is used in two forms of evaluation. The first is test prediction output which can be named the predicted knowledge. The second form of evaluation or testing is used to evaluate the output of the retention phase.

6. Retention phase. The aim of this retention phase is to provide a suitable retention solution to increase the customer lifetime to the maximum period of time registering with the service provider. During this phase there are three steps for reaching a retention solution which are:

  a) Classify churn attributes, b) available solutions (Solution Bundle), and c) assign solution.

7. Deployment phase. After testing the retention phase output then the discovered knowledge must be organized and presented in a generalized form that the use can use in future similar situations. For maintaining this phase it is important to monitor and maintain the available solutions by time.

IV.   Case Study Demonstration and result analysis

A data set for three months of usage with 5000 customers' records of an anonymous telecom company in Egypt is used for demonstrating and evaluating the proposed framework. 22 parameters are used in order to define five categories (behavioral, dissatisfactory,

profitability, cross-selling, demographic) of churners. Predicting the churners throughout comparative study between DT, SVM, and NN of this data set has been identified in a previously published paper [24]. Description of the dataset and testing set are shown in figure 3.

Figure 3a: data set classification   Figure 3b: testing set classification

Figure 3: dataset and testing set classification

The results of the predictive models in table 2 indicate that Neural Network outperforms Support Vector Machine but both outperform Decision Tree.

Table2: accuracy of predictive models

Technique	Accuracy	Error Rate	Predicted churners
DT	77.9% (779/1000)	22.1% (221/1000)	403/500 (80.6%)
SVM	83.6% (837/1000)	16.4% (164/1000)	416/500 (83.2%)
NN	84.1% (841/1000)	15.9% (159/1000)	421/500 (84.1%)

Based on the best results attained form the prediction techniques; the predicted 421 out of 500 churners are used by K-means algorithm to cluster them into 3 clusters. Table 3 shows the produced 3 clusters.

Table 3: Clustering Output for all attributes

Cluster #	Number of predicted churners (421)
Cluster 0	88
Cluster 1	208
Cluster 2	125

The output of the clustering step can be used for the knowledge usage step in order to assign a retention solution for a specific cluster or customer. The retention phase is based on the clustering analysis, association rules among the available attributes and weighted sum theory. The offered solutions used for retaining churners are 14 solutions. Table shows the solutions offered for each category of churners.

International Journal of Advances in Computer Science & Its Applications
Volume 6 : Issue 1   [ISSN 2250-3765]

Publication Date : 18 April, 2016

Table 4: Solutions offered for churners (S= solution)

Criteria	S_id	Solution Description	Class
Profitability	S1	500 minutes free	Class 0
	S2	250 minutes free	Class 1
	S3	125 Minutes free	Class 2
Cross-Selling	S4	Handset	Class 0&1
	S5	Free Mobile Internet	Class 0
	S6	300 minutes free + 100 sms	Class 2
Dissatisfactory	S7	200 minutes free + 50 sms	Class 0
	S8	100 minutes free + 25 sms	Class 1
	S9	50 minutes free + 15 sms	Class 2
Behavioral	S10	200 minutes free	Class 0
	S11	150 minutes free	Class 1
	S12	100 minutes free	Class 2
Demographic	S13	Distinct 3 numbers	Class 0&1
	S14	Weekend Offer (50 minutes free + 20 sms)	Class 2

The proposed CPRF presented in this paper uses the weighted sum method to rank the offers solutions in order to provide the best solution to the highest priority retention group. Table 5 shows the weights of the available solutions based on the experience of an expert with five years of experience working in an Egyptian telecom company in churn management department.

Table 5: Matrix for solutions with criteria and weights

Solution #	Profitability (0.4)	Behavior (0.2)	Dissatisfactory (0.2)	Cross-Selling (0.1)	Demographic (0.1)
	Relevancy				
S1	50	30	10	10	0
S2	30	30	20	10	10
S3	20	30	30	10	10
S4	30	20	20	30	0
S5	20	10	20	30	20
S6	40	20	20	20	0
S7	20	20	40	10	10
S8	20	20	30	10	20
S9	10	30	30	10	20
S10	20	40	20	10	10
S11	15	30	30	15	10
S12	10	30	30	15	15
S13	0	25	15	10	50
S14	0	20	20	10	50

Each solution is assigned a score based on the following formula

Score= ∑relevancy*weight

After computing the best solution to be assigned to the most important customers based on the high profitability and availability to churn. Solutions scores computed via using weighted sum score for each solution. Table 6 shows ranking for the available solutions and table 7 shows the result of assigning the best solution for each churner.

Table 6: Ranking retention solutions

Rank	S_id	Score	Rank	S_id	Score
1	S1	29	8	S8	21
2	S2	24	9	S9	20.5
3	S6	24	10	S10	19
4	S4	23	11	S11	19
5	S3	22	12	S12	19
6	S7	22	13	S13	14
7	S10	22	14	S14	14

Table 7: Solutions assignment regardless categories

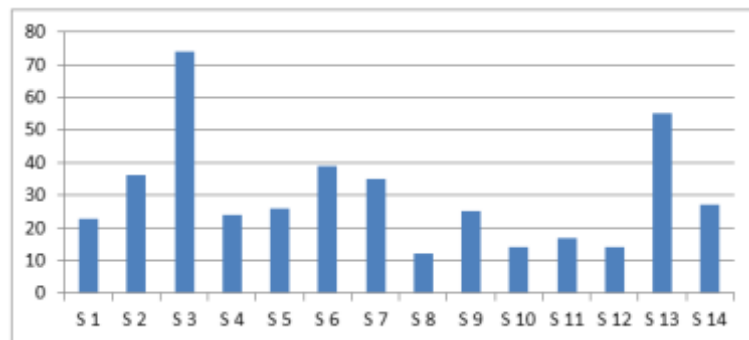| S_id	No. of customers	S_id	No. of customers
S 1	23	S 8	12
S 2	36	S 9	25
S 3	74	S 10	14
S 4	24	S 11	17
S 5	26	S 12	14
S 6	39	S 13	55
S 7	35	S 14	27
Total			421

Figure 4: Solutions assignment regardless categories

The results shown in table 7 and figure 4 indicate that solution 3 chosen to be the best for the maximum number of churners but solution 8 chosen to be the best for only 14 of churners which are 421 of predicted churners. The solutions assignment is accomplished regardless of the criteria but based on the churners clustering and the weighted sum method.

V. Conclusion

A lot of churn prediction models exist but the literature shows that most of them are focused on investigating the loyalty of customers throughout questionnaires to measure the customer satisfaction. So there is lack of studies regarding usage of computer techniques to find the best retention solutions. This paper proposes a churn prediction retention framework (CPRF) based on data mining and weighted sum theory. The proposed framework consists of 7 phases and other sub-phases.

The framework provides a methodology to use the extracted knowledge from the prediction phase to be used for assigning retention solution. Deploying the proposed framework into a BI application can help in enhancing the efficiency of CRM. But applying the proposed framework in real life BI environment is one of the glaring limitations of this paper. Although; it can help customer churn management department to easily predict and retain the expected future churners in many business areas. Regarding the future work, researchers should provide a model to detect whether or not a subscriber uses more than sim card to identify efficient retention solutions. Besides; there is no study till the submission of this thesis used Fuzzy Logic technique to predict churners in the mobile market.

VI.   References

[1]   V. Carlo, 2009. Business Intelligence: Data Mining and Optimization for Decision Making, John Wiley & Sons, Ltd. ISBN: 978-0-470-51138-1

[2]   H. Lee, Y. Lee, H. Cho, K. Im, Y. Kim, 2011. Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model," Decision Support Systems.

[3]   J. Hadden, A. Tiwari, R. Roy, D. Ruta, 2007. Computer Assisted Customer Churn Management : State-Of-The-Art and Future Trends. Computers & Operations Research, vol.34. pp.2902-2917

[4]   H. Kim, C. Yoon, 2004. Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. Telecommunications Policy. Vol. 28/ PP. 751-765.

[5]   D. Hoyer, J. Macinnis 2009. Consumer Behavior. 5th ed. South-Western

[6]   J. Urquizo, 2006. Improving and Monitoring Customer Retention. Microcredit Summit. http://www.microcreditsummit.org/papers/Workshops/31_Urquizo.pdf

[7]   M. Hassouna, 2012. Agent Based Modelling and Simulation: An Examination of Customer Retention in the UK Mobile Market. PhD thesis. Brunel University. UK

[8]   C. Rygielski, J. Wang, D. Yen, 2002. Data Mining Techniques for Customer Relationship Management, Technology in Society. Vol.24. pp.483-502.

[9]   K. Muata, O. Bryson, 2004. Evaluation of Decision Trees: A Multi Criteria Approach, Computers and Operational Research, vol.31. pp.1933-1945

[10]   S. Pal, and A. Ghosh, 2004. Soft Computing Data Mining, Information Sciences, 163, 1-3.

[11]   R. Palto, 2013. A Conceptual Framework of Customer Retention Strategy (CRS).   The Business & Management Review, Volume 3 Number 4. The International Conference on the Restructuring of the Global Economy (ROGE), London-UK

[12]   J. Zhi, Y. Xia, Y. Gan, 2012. Research and Design of Retention System in Telecommunications Companies. Pp.172–176

[13]   X. Li, Y. Huang, S. Li, Y. Zhang, 2011. Hybrid retention strategy formulation in telecom based on k-means clustering analysis. 2011 International Conference on E-Business and E-Government (ICEE), vol. 3. Pp.1–4. Doi:10.1109/ICEBEG.2011.5881956

[14]   M. Roberts-lombard, 2011. Customer retention through customer relationship management: The exploration of two-way communication and conflict handling, African Journal of Business Management Vol.5 (9). Pp. 3487-3496

[15]   E. Xevelonakis, 2005. Developing retention strategies based   on   customer   profitability   in telecommunications: An empirical study. Journal of Database   Marketing   &   Customer   Strategy Management. Vol. 12. Issue. 3. Pp.226–242. Doi:10.1057/palgrave.dbm.3240259

[16]   A. Haran, "Development of a framework to retain customers   through   customer   relationship management." (2005).

[17]   S. Aydin, and G. Ozer, 2005. The analysis of antecedents of customer loyalty in the Turkish mobile telecommunication market", European Journal of Marketing, vol. 39, no. 7, pp. 910-925.

[18]   W. Verbeke, D. Martens, C. Mues, B. Baesens, 2011. Building comprehensible customer churn prediction models   with   advanced   rule   induction techniques. Expert   Systems   with   Applications, Volume 38, Issue 3. Pp. 2354-2364

[19]   K. Silva, 2009. Customer Retention: With Special Reference to Telecommunication Industry in Sri Lanka, International Conference on Business and Information (July). Pp.1–15.

[20]   I. Khan, 2012. Impact of Customers Satisfaction And Customers   Retention   on   Customer   Loyalty, International Journal of Scientific & Technology Research Volume 1, Issue 2. ISSN 2277-8616 1(2)

[21]   A. Joachim, 2008. Customer service in the retention of mobile phone users in Nigeria, African Journal of Business Management Vol.2 (2), pp. 026-031.

[22]   K. Peighambari, 2007. Developing and Testing a Model for Explaining Customer Retention Formation: Case of Iranian Mobile Telecommunication Services. Master   Thesis. Faculty of Engineering, Tarbiat Modares University, Iran.

[23]   L. Ahn, S. Hana, Y. Lee, 2006. Customer churn analysis: Churn determinants and mediation effects of partial   defection   in   the   Korean   mobile telecommunications   service   industry. Telecommunications Policy. Vol. 30 No.: PP. 552-568

[24]   S. Essam, H. Yehia, K. Ayman, N. Mona, 2012. A Proposed Churn Prediction Model. International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622