INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT

IJRCM

A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories

Indexed & Listed at:

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A.

as well as in Open J-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)]

Registered & Listed at: Index Copernicus Publishers Panel, Poland

Circulated all over the world & Google has verified that scholars of more than 1500 Cities in 141 countries/territories are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

www.ijrcm.org.in

CONTENTS

CALL FOR MANUSCRIPTS

Weinvite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the area of Computer, Business, Finance, Marketing, Human Resource Management, General Management, Banking, Insurance, Corporate Governance and emerging paradigms in allied subjects like Accounting Education; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Monetary Policy; Portfolio & Security Analysis; Public Policy Economics; Real Estate; Regional Economics; Tax Accounting; Advertising & Promotion Management; Business Education; Management Information Systems (MIS); Business Law, Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labor Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; Public Administration; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism, Hospitality & Leisure; Transportation/Physical Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Digital Logic; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Multimedia; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic and Web Design. The above mentioned tracks are only indicative, and not exhaustive.

Anybody can submit the soft copy of his/her manuscript **anytime** in M.S. Word format after preparing the same as per our submission guidelines duly available on our website under the heading guidelines for submission, at the email address: infoijrcm@gmail.com.

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION:**

DATED: _____

THE EDITOR
IJRCM

Subject: **SUBMISSION OF MANUSCRIPT IN THE AREA OF .**

(e.g. Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript entitled '_____' for possible publication in your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if my/our manuscript is accepted, I/We agree to comply with the formalities as given on the website of the journal & you are free to publish our contribution in any of your journals.

NAME OF CORRESPONDING AUTHOR:
Designation:
Affiliation with full address, contact numbers & Pin Code:
Residential address with Pin Code:
Mobile Number (s):
Landline Number (s):
E-mail Address:
Alternate E-mail Address:

NOTES:

a) The whole manuscript is required to be in ***ONE MS WORD FILE*** only (pdf. version is liable to be rejected without any consideration), which will start from the covering letter, inside the manuscript.
b) The sender is required to mention the following in the **SUBJECT COLUMN** of the mail:
New Manuscript for Review in the area of (**Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/ Engineering/Mathematics/other, please specify)**
c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any specific message w.r.t. to the manuscript.
d) The total size of the file containing the manuscript is required to be below **500 KB**.
e) Abstract alone will not be considered for review, and the author is required to submit the complete manuscript in the first instance.
f) The journal gives acknowledgement w.r.t. the receipt of every email and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending separate mail to the journal.

2. **MANUSCRIPT TITLE:** The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.

3. **AUTHOR NAME (S) & AFFILIATIONS:** The author (s) **full name**, **designation**, **affiliation** (s), **address**, **mobile/landline numbers**, and **email/alternate email address** should be in italic & 11-point Calibri Font. It must be centered underneath the title.

4. **ABSTRACT:** Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para. Abbreviations must be mentioned in full.

5. **KEYWORDS:** Abstract must be followed by a list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stops at the end.

6. **MANUSCRIPT:** Manuscript must be in ***BRITISH ENGLISH*** prepared on a standard A4 size ***PORTRAIT SETTING PAPER***. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of every page. It should be free from grammatical, spelling and punctuation errors and must be thoroughly edited.

7. **HEADINGS:** All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.

8. **SUB-HEADINGS:** All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.

9. **MAIN TEXT:** The main text should follow the following sequence:

INTRODUCTION

REVIEW OF LITERATURE

NEED/IMPORTANCE OF THE STUDY

STATEMENT OF THE PROBLEM

OBJECTIVES

HYPOTHESES

RESEARCH METHODOLOGY

RESULTS & DISCUSSION

FINDINGS

RECOMMENDATIONS/SUGGESTIONS

CONCLUSIONS

SCOPE FOR FURTHER RESEARCH

ACKNOWLEDGMENTS

REFERENCES

APPENDIX/ANNEXURE

It should be in a 8 point Calibri Font, single spaced and justified. The manuscript should preferably not exceed ***5000 WORDS***.

10. **FIGURES &TABLES:** These should be simple, crystal clear, centered, separately numbered &self explained, and **titles must be above the table/figure**. **Sources of data should be mentioned below the table/figure**. It should be ensured that the tables/figures are referred to from the main text.

11. **EQUATIONS:**These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.

12. **REFERENCES:** The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow **Harvard Style of Referencing**. The author (s) are supposed to follow the references as per the following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s)** for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parentheses.
- The location of endnotes within the text should be indicated by superscript numbers.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:

BOOKS

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–22 June.

UNPUBLISHED DISSERTATIONS AND THESES

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, KurukshetraUniversity, Kurukshetra.

ONLINE RESOURCES

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITES

- Garg, Bhavet (2011): Towards a New Natural Gas Policy, Political Weekly, Viewed on January 01, 2012 http://epw.in/user/viewabstract.jsp

APPLYING AND EVALUATING DATA MINING TECHNIQUES TO PREDICT CUSTOMER ATTRITION: A SURVEY

AFAQ ALAM KHAN
ASST. PROFESSOR
DEPARTMENT OF INFORMATION TECHNOLOGY
CENTRAL UNIVERSITY OF KASHMIR
SRINAGAR

NASIR HUSSAIN
ASST. PROFESSOR
DEPARTMENT OF INFORMATION TECHNOLOGY
CENTRAL UNIVERSITY OF KASHMIR
SRINAGAR

PARVEZ ABDULLAH KHAN
ASST. PROFESSOR
DEPARTMENT OF MANAGEMENT STUDIES
BABA GHULAM SHAH BADSHAH UNIVERSITY
RAJOURI

ABSTRACT

Churn prediction is the concept of identifying those customers who are intending to move their custom to a competing service provider. This phenomenon has been magnified by electronic commerce. The internet channel returns control and power to customers who are no longer confined to the decision of single company. The outcomes are increase in customer power (Minguel, 2005) and competition exacerbation. Customer empowerment is likely to persist and amplify customer attrition issue. This paper focuses on introducing the phenomenon of customer attrition, reason for customer attrition and various data mining techniques which the author has encountered in literature. These techniques have been successfully applied to predict customer churn in different domains. Most common tools and measures used for evaluating the accuracy of different data mining techniques is another important issue which has been addressed in this paper.

KEYWORDS
Customer churn, classification, data mining, ISP, Prediction.

1. CUSTOMER ATTRITION

With the emergence of electronic commerce, customers can more easily inquire about the market opportunities. They become more demanding and tend to switch from current/previous service provider to another service provider. This gives birth to the notion of churn, the movement of customers from provider to provider in search of better and cheaper products and services. Churn or customer attrition is defined as "the annual turnover of market base" (Minguel, 2005).

Geppert (2002) defines churn as the gross rate of customer loss during a given period. Churn can be shown as follows:

Monthly Churn = (C0 + A1 - C1) / C0 *(1.1)*

Where:

C0 = Number of customers at the start of the month

C1 = Number of customers at the end of the month

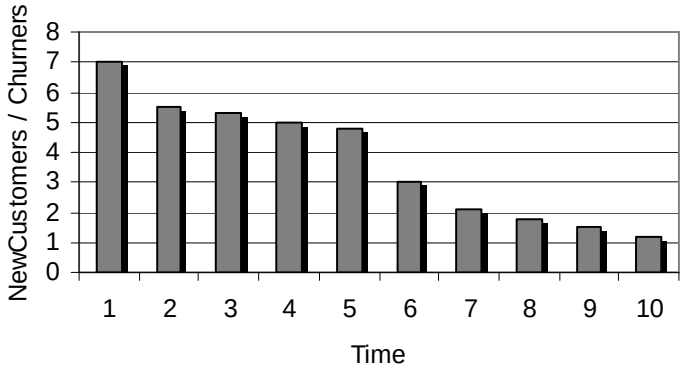
A1 = Gross new customers during the month

Churn prediction is the concept of identifying those customers who are intending to move their custom to a competing service provider. Once identified, these customers can be targeted with proactive marketing campaigns for retention efforts. Churning customers can be divided into two main groups, voluntary and non-voluntary churners (Daskalaki, 2003). Non-voluntary churners are the easiest to identify, as these are the customers who have had their service withdrawn by the company. There are several reasons why company could revoke a customer's service, including abuse of service and non-payment of service. Voluntary churn is more difficult to determine, because this type of churn occurs when a customer makes a conscious decision to terminate his/her service with the provider. Voluntary churn can be sub divided into two main categories, incidental churn and deliberate churn. Incidental churn happens when changes in circumstances prevent the customer from further requiring the provided service. Example of incidental churn includes changes in the customer's financial circumstances, so that the customer can no longer afford the service, or a move to different geographical location where the company's service is unavailable. Incidental churn usually only explains a small percentage of company's voluntary churn. Deliberate churn, also the major concern of this study, is the problem that most churn management solutions tries to battle (Gappert, 2002). Reasons that could lead to customer's deliberate churn are discussed in section 2.

This phenomenon has been magnified by electronic commerce. The internet channel returns control and power to customers who are no longer confined to the decision of single company. The outcomes are increase in customer power (Minguel, 2005) and competition exacerbation. Customer empowerment is likely to persist and amplify customer attrition issue.

The impact of churn is correlated with the industry lifecycle. When the industry is in the growth phase of its life cycle, sales increase exponentially;

FIGURE 1.1: CHURN RATE EVOLUTION (Minguel, 2005)

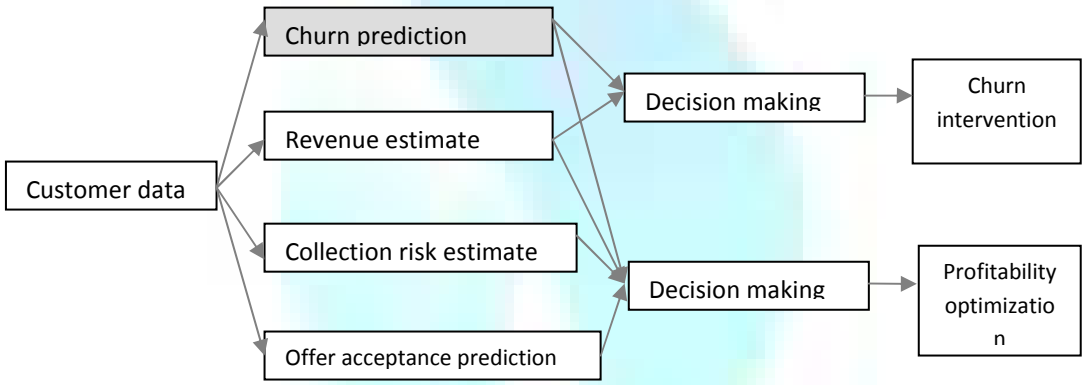


The number of new customers largely exceeds the number of churners. Companies aim at getting more and more new customers. Nevertheless, the ratio (new customers/churners) tends one over time (Figure 1.1). The impact of churn becomes then markedly more sensitive. For product/services in the maturity phase of their lifecycle, companies put the focus on the churn rate reduction.

As shown in Figure 1.2, Yan et.al (2004) state that predicting customer churn is a component in the decision framework for retaining customers and maximizing profitability. Yan et.al (2004) further state that customer data can help predict the reasons customers switch companies, the revenues generated in a specified period, revenue collection risk, and the response to a marketing offer.

Companies can use these probability and revenue estimates in a decision theoretic framework to determine a churn intervention strategy and a profitability optimization strategy.

FIGURE 1.2: THE FRAMEWORK FOR CUSTOMER RETENTION AND PROFITABILITY MAXIMIZATION (Yan et.al, 2004)



There are two basic approaches of managing customer churn. Untargeted approaches rely on superior product and mass advertising to increase brand loyalty and retain customers. Targeted approaches rely on identifying customers who are likely to churn, and then either providing them with a direct incentive or customizing a service plan to stay. There are two types of targeted approaches: reactive and proactive. With the reactive approach, the firm waits until the customer contacts the firm to cancel his or her account. The firm then offers the customer an incentive to stay. With the proactive approach, the firm tries to identify in advance customers who are likely to churn at some later date. The firm then targets these customers with special programs or incentives to forestall the customer from churning (Burez, 2006). Churn management efforts should not focus across the entire customer base because firstly not all customers are worth retaining, and secondly customer retention costs money; attempting to retain customers that have no intention of churning will waste resource. Targeted proactive programs have potential advantages of lower incentive costs.

2. REASONS OF CHURN

In an intensely competitive environment like that of mobile services or internet services, customers receive numerous incentives to switch and encounter numerous disincentives to stay.

- Price. Particularly in the wireless and long-distance markets, carriers often offer pricing promotions, such as relatively low monthly fees, high-volume offerings (fixed number of minutes at a reasonable fee per month), and low rates per-minute. These price incentives can provide residential customers, in particular, with powerful incentives to change carriers.
- Service quality. Lack of connection capabilities or quality in places where the customer requires service can cause customers to abandon their current carrier in favor of one with broader reach or a more robust network. Frequent disconnections from their service providers may also increase the user's frustration.
- Fraud. Customers may attempt to "game the system" by generating high usage volumes and avoiding payment by constantly churning to the next competitor.
- Lack of carrier responsiveness. Slow or no response to customer complaints is a sure path to a customer relations disaster. Broken promises, long hold times when the customer reports problems, and multiple complaints related to the same issue are sure to lead to customer churn.
- Brand disloyalty (or loyalty to another). Brand issues may arise due to service or other issues experienced over time, mergers or acquisitions involving the incumbent carrier, or entry into the market of another carrier with strong brand recognition and reputation. Marginal brand loyalty can often be overcome by competitors' incentives.
- Privacy concerns. Consumers have an increasing awareness that companies they deal with have a lot of information about them, including their spending habits, personal financial information, health information, and the like. Breaking of privacy promises, publicized privacy problems, telemarketing, and other

issues are causing many customers to consider their personal privacy as an asset and they are holding their service providers responsible for keeping privacy promises.

- Lack of features. Customers may switch carriers for features not provided by their current carrier. This might include the inability of a particular carrier to be the "one- stop shop" for all the customer's communications needs.
- New technology or product introduced by competitors. New technologies —such as high-speed data or bundled high-value service offerings—create significant opportunities for carriers to entice competitors' customers to switch.
- New competitors enter the market. The mere existence of viable competitors to the incumbent carrier may cause certain disloyal customers to churn. Further, as competitors enter new markets, they often offer short- or long-term incentives to new subscribers to build market share.
- Billing or service disputes. Billing errors, incorrectly applied payments, and disputes about service disruptions can cause customers to switch carriers. Depending on the situations, such churn may be avoidable.
- Filtering, depending upon the government rules and regulations ISPs may be asked to block some content, which may cause inconvience to some users. Especially when filtering is not accurate and some sites are blocked by mistake.

3. APPLYING DATA MINING TO DEAL WITH CUSTOMER CHURN:

To solve a business problem like customer churn using data mining, that problem is needed to be translated in to one of the KDD functions. These functions include Classification, Prediction, Estimation, Association, Clustering, etc. Classification, Prediction and Estimation are examples of directed data mining. Association and Clustering are examples of undirected data mining. In directed data mining there is always a target variable- something to be classified, estimated or predicted. In undirected data mining, there is no target variable. The data mining task is to find overall patterns that are not tied to any one variable. The most common form of undirected data mining is clustering, which finds groups of similar records without any instructions about which variable should be considered as most important. In the machine learning literature, directed data mining is called supervised learning and undirected data mining is called unsupervised learning (Berry, 2004). Churn prediction being a classification task (Wei et.al, 2002) it will be discussed next section including its difference with prediction.

3-1) CLASSIFICATION AND PREDICTION

Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. Such analysis can help provide us with a better understanding of the data at large. Whereas classification predicts categorical (discrete, unordered) labels, prediction models continuous-valued functions (Han, 2003).

CLASSIFICATION is a two-step process, In the first step, a classifier is built describing a predetermined set of data classes or concepts. This is the learning step (or training phase), where a classification algorithm builds the classifier by learning from a training set made up of database tuples and their associated class labels. A tuple, X, is represented by an n-dimensional attribute vector, X = (x1, x2, …., xn), depicting n measurements made on the tuple from n database attributes, respectively, A1,A2,….,An. Each tuple, X, is assumed to belong to a predefined class as determined by another database attribute called the class label attribute. The class label attribute is discrete-valued and unordered. It is categorical in that each value serves as a category or class. The individual tuples making up the training set are referred to as training tuples and are selected from the database under analysis.

This first step of the classification process can also be viewed as the learning of a mapping or function, y = f(X), that can predict the associated class label y of a given tuple X. In this view, we wish to learn a mapping or function that separates the data classes. Typically, this mapping is represented in the form of classification rules, decision trees, or mathematical formulae.

In the second step, the model is used for classification. First, the predictive accuracy of the classifier is estimated. If we were to use the training set to measure the accuracy of the classifier, this estimate would likely be optimistic since the classifier tends to overfit the data (that is, during learning it may incorporate some particular anomalies of the training data that are not present in the general data set overall). Therefore, a test set is used, made up of test tuples and their associated class labels. These tuples are randomly selected from the general data set. They are independent of the training tuples, meaning that they are not used to construct the classifier. The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier. The associated class label of each test tuple is compared with the learned classifier's class prediction for that tuple. If the accuracy of the classifier is considered acceptable, the classifier can be used to classify future data tuples for which the class label is not known. (Such data are also referred to in the machine learning literature as "unknown" or "previously unseen" data.) (Han, 2003).

PREDICTION is a two-step process, similar to that of data classification. However, for prediction, we lose the terminology of "class label attribute" since the attribute for which values are being predicted is continuous-valued (ordered), rather than categorical (discrete-valued and unordered). The attribute can be referred to, simply, as the predicted attribute. Prediction can also be viewed as a mapping or function, y = f(X), where X is the input (e.g., a tuple describing a loan applicant) and the output y is a continuous or ordered value (such as the predicted amount that the bank can safely loan the applicant). That is, we wish to learn a mapping or function that models the relationship between X and y.

Prediction and classification also differ in the methods that are used to build their respective models. As with classification, the training set used to build a predictor should not be used to assess its accuracy. An independent test set should be used instead. The accuracy of a predictor is estimated by computing an error based on the difference between the predicted value and the actual known value of y for each of the test tuples, X (Han, 2003).

4. CUSTOMER CHURN IN LITERATURE

Customer attrition is an important issue for any company and is easiest to define in subscription based businesses, and partly for that reason, churn modeling is most popular in theses businesses (Berry, 2004). Long-distance companies, Mobile phone service providers, Insurance companies, Cable companies (Pay-TV) (Burez et.al, 2006), financial service companies, Internet service providers, newspapers, magazines, and some retailers all share subscription model where customers have a formal, contractual relationship which must be explicitly ended.

4-1. CUSTOMER CHURN IN MOBILE SERVICE PROVIDERS

As deregulation, new technologies, and new competition have opened up telecommunication industry, the telecommunication service market has become more competitive then ever. And in this strongly competitive and broadly liberalized mobile telecommunication industry, customer churn has turned into very serious issue. Many subscribers frequently churn from one provider to another in search of better rates/service or for the benefits of signing up with new carrier (e.g., such as receiving the latest cellular phone). Low signup fees particularly for prepaid mobile services also encourage customers to churn. Wireless local number portability (WLNP) or the ability to change mobile carriers and keep mobile phone number also poses a big challenge to the already reeling telecommunication company.

It is estimated that the average churn rate for the mobile telecommunications is 2.2% per month. i.e., about 27% of given carrier's subscriber are lost each year (Yan et.al, 2004), making it essential to develop an effective churn reduction method. The cost of acquisition of new mobile service subscriber is estimated to be from $300 to $ 600. However, the cost of retaining an existing subscriber is generally much lower than that (Wei et.al, 2002).

Wei et.al (2002) used the call pattern changes and contractual data and developed a data mining based churn prediction technique that identify potential churners at the contract level. In their study they used the decision tree approach as the basis for the development of their technique for mobile telecommunication company in Taiwan. In an another attempt to identify the determinants of subscriber churn and customer loyalty in Korean mobile telephony market; Kim et.al (2004) used binomial logit model based on the survey of 973 mobile users .

4-2. CUSTOMER CHURN IN BANKING AND INSURANCE

Over the past two decades, the financial markets have become more competitive due to the mature nature of the sector on the one hand and deregulation on the other, resulting in diminishing profit margins and blurring distinctions between banks, insurers and brokerage firms. Hence, now a day a small number of large institutions offering a wide set of services dominate the financial-service industry. These developments stimulated bank and Insurance companies to

implement customer relationship management (CRM). Under this intensive competitive pressure, companies realize the importance of retaining, their current customers. The substantive relevance of attrition modeling comes from the fact that a bank is able to increase its profits by 85% due to a 5% improvement in the retention rate (Prinzie et.al, 2005).

Bart Lariviere et.al (2004) studied the defectors of saving and investment (SI) customers of a large Belgian financial service provider. Their study involves the use of Kplan-Meier estimator to gain insight into the timing of the SI churn event. A multi dimensional probit model and a proportional hazard model are performed to find the most convenient products to cross sell in terms of customer preferences and the likelihood to lower the customer defection proness respectively.

In order to establish a system that can provide warning signs to network banking services soon before they are losing valuable customers, Chaing et.al. (2003) took association rules in use and proposed an efficient algorithm called goal- oriented sequential pattern, which can find out behavior patterns of loosing customers or clues before they stop using some products.

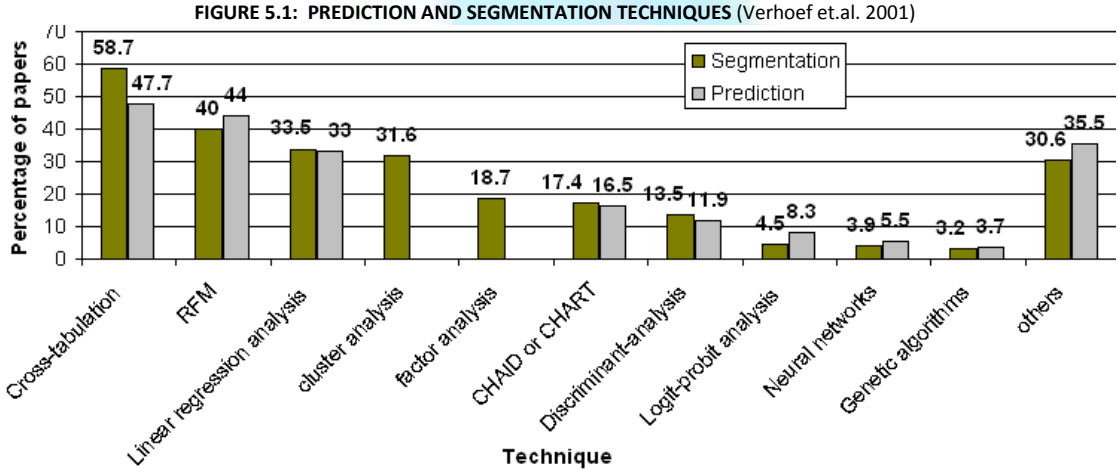
4-3. CUSTOMER CHURN IN ISP

Internet popularity is growing at impressive rate. Sooner or later, every customer comes face to face with a decision of choosing an ISP. ISPs endure a five times higher churn rate cumulating to 10% monthly (Minguel, 2005). It is also said that nearly half of all Internet subscribers leave their provider every year (Au et.al, 2003). IP network metering, which has lately pick up steam, is the key enabler of getting the fundamental information of churn prediction. Shen-Tun Li et.al. (2005) applied the well known data mining methodology CRISP-DM to investigate network usage behaviors of the ISP subscribers in Taiwan. They used Attribute-oriented Induction (AOI) method for discovering characteristics and discrimination knowledge of ISP customers from the ISP traffic data. An empirically tested model developed by Jyh-Shen Chiou (2004) examines the antecedents of consumer loyalty toward ISPs. In choosing an ISP, pricing was and is a main differentiating factor in mind of customers; however, service quality has emerged as a major concern among users lately. Management of ISP has discovered that service quality is important not only for attracting new customers, but more importantly for retaining existing customers who may otherwise be lured away by lower fees. Gray Madden et.al (1999) developed a probability model for ISP subscriber churn. Their model relates the probability of subscribers churn to various service attributes and subscriber characteristics. Khan et.al (2010) used the subscribers' demographic, billing and usage date to find the best predictors in one hand and compared the performance of different data mining techniques on the other. They found that best predictors lies among usage data and neural network showed better performance than decision trees.

5. TECHNIQUES USED TO DEAL WITH CHURN

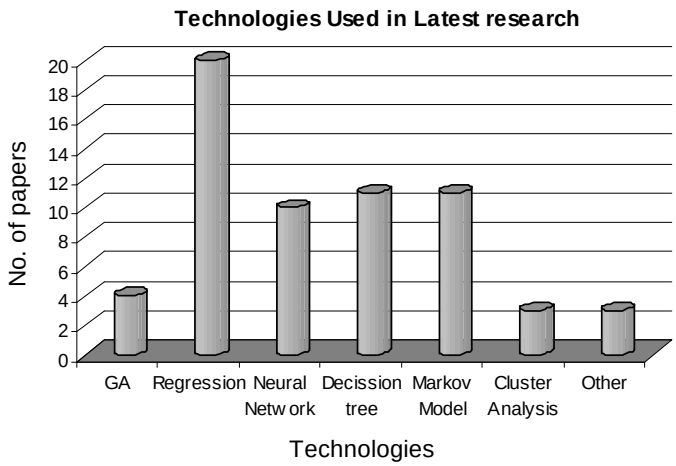
This section provides a discussion on the various predictive modeling techniques and provides an overview of the research gap area. Much work has been done in the area of CRM and predicting customer behavior, although it appears that to date customer churn management has not received the attention that it requires. Figure 5.1 shows a chart of the known predictive and segmentation methods, compiled by Verhoef and Donkers (2001). It can be observed from Figure 5.1 that cross-tabulation has been the most popular segmentation and predictive modeling technique in the past. The next column shows RFM as segmentation and predictive modeling technique. Hadden et.al (2005) state that RFM defines the variables for use with segmentation and prediction modeling techniques rather than a technology of its. Methods such as waiting RFM variables have been reported in the literature, and weighting the variables provides better results.

FIGURE 5.1: PREDICTION AND SEGMENTATION TECHNIQUES (Verhoef et.al. 2001)



Linear regression is another popular technique, for performing both segmentation and prediction, while cluster analysis and factor analysis have been shown specifically for segmentation. According to Verhoef et.at (2001), the three main technologies for prediction, CART, Logit analysis and neural networks have been least used methods. However, the figure shows that CART has been preferred over neural networks and logit regression.

FIGURE 5.2: KDD TECHNIQUES USED IN LATEST RESEARCH (Hadden, et. al., 2005)



Based on a review of literature related to churn management, Hadden et.al, (2005) compiled Figure 5.2. The research conducted by Hadden et.al, (2005) is focused mainly on publications during the last five years. Therefore it updates the finding of Verhoef et.al, (2001). Cross-tabulation is an older modeling technique, which has been overtaken by more advanced and accurate methods. Furthermore, the difference between Figure 5.1 and 5.2 are also because the focus of Figure 5.1 is on segmentation and prediction techniques in general, and focus of figure 5.2 is on technologies for predicting customer behavior. One of the major differences is that regression analysis has taken the lead for the preferred method, following by decision trees and Markov models, and then neural networks. This means that four of the least used methods reported by Figure 5.1 have become four of the most popular as reported by Figure 5.2. Following the charts, Table 5.1 is presented, which offers a reference to which research papers are connected with which technology.

TABLE 5.1: REFERENCE OF PAPERS AND THE TECHNOLOGIES

Reference	Regression	Neural Network	Decision Tree	Markov model	Discrimenant analysis	GA	Other
Au, W.-H., (2003)		*	*				
Hwang, (2004)	*	*	*				
Baesens, (2004)	*			*			
Ng and Liu, (2001)			*				
Verhoef & Donkers, (2001)	*						
Ven Den Poel, (2003)	*						
Madden & Savage, (1999)	*						
Wei & Chiu, (2002)			*				
Van Den Poel & Lariviere, (2003)			*				
Chae, (2001)	*		*				
Au, T., (2003)	*	*	*				
Daskalki, (2003)		*	*		*		
Malhotra, (2003)		*	*		*		
Chiang, (2003)		*				*	
Lariviere, (2004)	*						
Hsieh, (2004)		*					
Hung, (2003)		*	*				
Kim , (2004)	*						
Yan, (2004)	*						
Burez, (2006)				*			

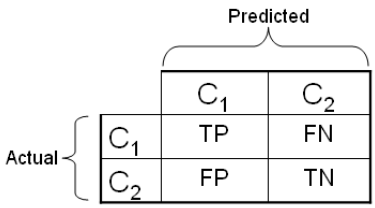
Support Vector Machine is another technique which has lately picked up and is being extensively used for churn prediction.

A Support Vector Machine (SVM) is an algorithm for the classification of both linear and nonlinear data. It transforms the original data in a higher dimension, from where it can find a hyperplane for separation of the data using essential training tuples called support vectors. Detailed description of all the classification techniques is out of the scope of this paper. However, next section presents how we can calculate the accuracy of different models applied for churn prediction.

1. EVALUATING DIFFERENT TECHNIQUES

The confusion matrix is a useful tool for analyzing how well our classifier can recognize tuples of different classes. A confusion matrix for two classes is shown in fig.

FIG 6.1: CONFUSION MATRIX FOR TWO CLASSES



As an example, if we consider C1 as a class of churners and C2 as a class of non churners, then TP (true positive) represents those who were churners and are correctly classified as churners, FP (false positive) represent those who where non churners but our classification technique has classified them as churners. Similarly, FN (False negative) represent those who were churners but are classified as non churners and TN (True negative) represent those who are non churners and are correctly classified as non churners.

The following meaningful measures can be extracted (Bradley (1997)):

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Sensitivity = TP / TP + FN

Specificity = TN / (TN + FP)

There are several methods documented for validating a customer churn model. Some popular methods are discussed below:

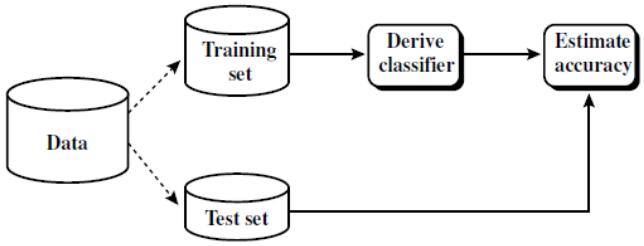
• Cross-fold validation: Hwang et.al, performed validation by creating a 70/30 divide of the data. The 70% divide created the training set, and the 30% divide created the validation set. Cross-fold validation is based on the principle of using the available for both training and validation. Several cross-validation methods have been proposed in the literature, two examples follow:

◦ V-fold cross validation—the learning set is randomly partitioned into limited datasets of equal size. Each set is then used as a validation set.

◦ Monte Carlo cross validation—the learning set is repeatedly divided into two random sets for training and validation.

• Cross-fold validation is most suitable in those cases in which there is a scarcity of data.

FIG 6.2: ESTIMATING ACCURACY WITH HOLDOUT METHOD

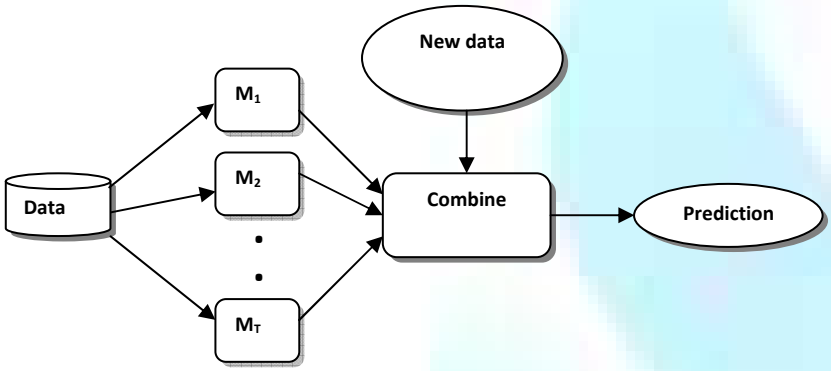


•The holdout method shown in fig 6.2, the given data are randomly partitioned into two independent sets, a training set and a test set. Typically, two thirds of the data are allocated to the training set, and the remaining one third is allocated to the test set. The training set is used to derive the model, whose accuracy is estimated with the test set.

7. INCREASING THE ACCURACY

There have been numerous comparisons of the different classification and prediction methods, and the matter remains a research topic. No single method has been found to be superior over all others for all data sets. Issues such as accuracy, training time, robustness, interpretability, and scalability must be considered and can involve trade-offs, further complicating the quest for an overall superior method.

FIG 7.1: INCREASING MODEL ACCURACY (Han, 2003)



Empirical studies show that the accuracies of many algorithms are sufficiently similar that their differences are statistically insignificant, while training times may differ substantially. However, there are strategies to increase the accuracy of prediction or classification. Two such techniques are Bagging and Boosting; each generate a set of classification or prediction models, M1, M2,……,Mk. Voting strategies are used to combine the predictions for a given unknown tuple.

CONCLUSION

Customer churn has been identified as one of the serious business issues and the mitigation of the same has been the center of attention for academicians and researchers from quite some time. Without going in to the details of different data mining techniques, which has been taken use of, for the minimization of customer churn, this paper first of all introduced the notion of customer churn in detail and highlighted different possible reasons which give birth to the churn. Different domains, which are severely affected with customer churn have been identified as appeared in the literature. However, different indexes and parameters for measuring the most affected industry are yet to be investigated. This paper further introduced the most common tools and measures used for evaluating the accuracy of different data mining techniques.

REFERENCES

1. Au, T., Li, S., and Ma, G., (2003). "Applying and Evaluating to Predict Customer Attrition Using Data Mining Techniques." Journal of Comparative International management 6(1).
2. Au, W.-H., Chan, K. C. C., and Yao, X., (2003). "A Novel Evolutionary DataMining Algorithm With Application to Churn Prediction." IEEE Transactions on Evolutionary Computation 7(6): 532-545.
3. Baesens, B., Verstraeten, G., Van den Poel, D., Egmont-Peterson, M., Van Kenhove, P., and Vanthienen, J., (2004). "Bayesian network classifiers for identifying the slope of the customer lifecycle of long life customers." European Journal of Operational Research 165: 508-523.
4. Berry, J. A., and Linoff, G. S., (2004). Data Mining Techniques For Marketing, Sales, and Customer Relationship Management. USA, Wiley.
5. Burez, J., and Van den Poel, D., (2006). "CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services." Expert Systems With Applications 101: 512-524.
6. Chae, Y. M., Hee Ho, S., Won Cho, K., Lee, D. H., and Ha Ji, S., (2001). "Data mining approach to policy analysis in a health insurance domain." International journal for Medical Informatics 62: 103-111.
7. Chiang, D.-A., Wang Y.-F, and Lee, S.,-L., (2003). "Goal-oriented sequential pattern for network banking churn analysis." Expert System with Applications 25: 293-302.
8. Chiou, Jyh-Shen., (2003). "The antecedents of customers' loyalty toward Internet Service provider". Information & Management 41: 685-695
9. Daskalaki, S., Kopanas, I., Goudara, M., and Avouris, N., (2002). "Data mining for decision support on customer insolvency in telecommunication business." European Journal of Operational research 145(10): 239-255.
10. Gappert, C., (2002). "Customer Churn in the Communications Industry." A KPMG LLP white paper, U.S member of KPMG International.
11. Hadden, J., Tiwari, A., Roy, R., and Ruta, D., (2005). "Computer assisted customer churn management: Stat-of-the-art and future trends." Computers & Operations Research 10(2): 139-158.
12. Hadden, J., Tiwari, A., Roy, R., and Ruta, D., (2005). "Computer assisted customer churn management: Stat-of-the-art and future trends." Computers & Operations Research 10(2): 139-158.
13. Han, J., Kamber, M., (2003). Data Mining: Concepts and Techniques, Morgan Kaufmann.
14. Hsieh, Nan-Chen., (2004). "An integrated data mining and behavioral scoring model for analyzing bank customers." Expert Systems with Applications 27: 623-633.

15. Hung, S., -Y., and Wang, H.-Y., (2004). Applying Data Mining to Telecom Churn Management. Proceedings of third annual conference, Pacific Asia Conference on Information Systems, Taiwan.
16. Hwang, H., (2004). "An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunications industry." Expert Systems with Applications 26: 103-120.
17. Kantardzic, M. (2003). Data Mining: Concepts, Models, Methods, and Algorithms, John Wiley & Sons.
18. Khan, A. A., Jamwal, S., and Sepehri, M. M., (2010) "Applying Data Mining to Customer Churn Prediction in an Internet Service Provider" International Journal of Computer Applications 9(7):8–14.
19. Kim, H.-S., and Yoon, C.-H., (2004). "Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market." Telecommunication Policy 28(6): 751-765.
20. Lariviere, B., and Van den Poel, D., (2004). "Investigating the role of product features in preventing customer churn by using survival analysis and choice modeling: The case of financial services." Expert Systems with Applications 27(2): 277-285.
21. Li, S.-T., Shue, L.-Y., and Lee, S.-F., (2005). "Enabling Customer Relationship Management in ISP Services Through Mining Usage Patterns." Expert System with Applications 100(23): 1-12.
22. Madden, G., Scott J. S., and Cobel-Neal, G., (1999). "Subscriber Churn in Australian ISP Market." Information Economics and policy 11(12): 195-207.
23. Malhotra, R., Malhotra, D. K., (2003). "Evaluating consumer loans using neural networks." International journal of management science 31: 83-96.
24. McKay, H., (2005). "Development of contractual relationship between an ISP and its customers - Is a fairer deal in sight?" Computer Law and Security Report 21: 216-225.
25. Minguel, A. P. M. L., (2005). "Measuring the impact of data mining on churn management." Internet Research: Electronic Network Applications and Policy 11(5): 375-387.
26. Mitra, S., Acharya, T., (2003). Data Mining: Multimedia, Soft Computing, and Bioinformatics. USA, John Wiley & Sons.
27. Ng, K., Liu, H., (2001). "Customer Retention Via Data Mining." Issues on the Application of data Mining: 569-590.
28. Prinzie, A., Van den Poel, D., (2005). " Incorporating sequential information into traditional classification models by using an element/position-sensitive SAM". Decision Support Systems.
29. Van den Poel, D., (2003). "Predicting Mail-Order Repeat Buying: Which Variables Matter?" Tijdschrift voor Economie and management 48(3): 371-403.
30. Van Den Poel, D., Lariviere, B., (2003). "Customer attrition analysis for financial services using proportional hazard models." European Journal of Operational Research 157: 196-217.
31. Verhoef, P. C., Donkers, B., (2001). "Predicting customer potential value an application in the insurance industry." Decision Support System 32: 189-199.
32. Wei, C.-P., and Chiu, I.-T., (2002). "Turning telecommunications call details to churn prediction: a data mining approach." Expert Systems with Applications 23(10): 103-112.
33. Yan, L., Wolniewicz, R. H., and Dodier, R., (2004). "Predicting Customer Behavior in Telecommunications." IEEE Intelligent Systems: 50-58.

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Computer Application and Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you to supply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mail **infoijrcm@gmail.com** for further improvements in the interest of research.

If you have any queries please feel free to contact us on our E-mail **infoijrcm@gmail.com**.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-

Co-ordinator

ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

Our Other Journals