Volume 4, Issue 12, December 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Survey on Profit Maximizing Metric for Measuring Classification Performance of Customer Churn Prediction Models

Niketa V Patil, Dr. A M Dixit
Computer Engineering, PVPIT,
Pune, India

Abstract: Customer churn prediction is one of the problems that most concern to businesses today. Predictive models can be developed for identifying future churners. This paper surveys the commonly used data mining techniques to identify customer churn patterns. The recent literature in the area of predictive data mining techniques in customer churn behaviour is reviewed and a discussion on the future research directions is offered and an introduction to ROC graphs and as a guide for using them in research.

Keywords: Customer Churn, Customer relationship management, ROC, Data Mining, Data mining technique,

I. INTRODUCTION

Data volume has been growing at a tremendous pace over the last two decades due to advancements in information technology(IT). At the same time there has been enormous development in data mining. Most of new methods and techniques have been added to process data and gather information. The data gathered from any source is raw data. In which the valuable information is hidden[3]. Data mining can be defined as the process of extracting valuable information from data. Data mining techniques have been successfully applied in many different domains and fields. Customer retention is one of the fundamental aspects of Customer Relationship Management, especially within the current economic environment, since it is more profitable to keep existing customers than attract new customer.

A small improvement in customer retention can produce an increase in the profit. The early detection of future churners is one of the Customer relationship management strategies. Predictive models provide us with a numerical measure that assigns to each client their propensity to churn in terms of probability. The higher the propensity value assigned to a customer, the greater their inclination to leave the firm. This information can be used by the company to develop marketing campaigns aimed at customer retention. To build such predictive models, several statistical strategies for classification technique can be used. After being built, predictive models have to be validated. This can be done in terms of different criteria, for instance: accuracy, speed, robustness, interpretability and ease of use. It is common to develop several models using different statistical strategies to compare them and select the most appropriate classification system for a specific problems. Verbeke et al provide an overview of the literature on the use of data mining techniques for customer churn prediction modelling. They show the characteristics of the assessed datasets, the different applied modelling techniques and the validation and the evaluation of the results. Different evaluation parameters are used in the considered studies. Most of them being mainly based on aspects related to the accuracy of the model. The metrics that were more frequently used are percentage of correctly classified (PCC)[1], area under curve (AUC).

Regarding classification techniques, some authors focus at the individual classifiers, such as logistic regression, neural networks and classification trees. Moreover, thanks to the improvement in computer hardware. Other techniques have been recently developed as a combination of other individual classifiers. Some examples of these techniques are the Random forest (based on the bagging) and AdaBoost (based on the boosting). Despite of this diversity, we didn't find works which evaluated the accuracy against other parameters such as speed, robustness, interpretability and the ease of use. Those other parameters are also important, as a classification model has not only to be accurate but also interpretable, usable and implementable by the final users.

In fact, some authors conclude that the different predictive techniques happened to show similar accuracy levels on their data, so other discriminate parameters could be used. Buckinx et al. work on real data from a retail company. They use logistic regression, neural networks and random forest as classification techniques and the performance is assessed through the percentage of correctly classified and AUC. Besides the standard measures of accuracy, we the results of a study on a financial services company in Belgium. In this case, a framework is provided for evaluating churner classification techniques based on the profit loss incurred by a misclassification cost, considered from a customer lifetime value perspective. At the same way,Verbeke et al. propose a profit centric performance measure. They use the idea of maximum profit to evaluate customer churn prediction model in the telecommunication sector.

As more and more methods are elaborated, the need for adequate performance measures has become more important than ever before. There has been a lot of attention for the receiver operating characteristic (ROC)[2] curve, which is a

© 2014, IJARCSSE All Rights Reserved

graphical representation of the classification performance for varying thresholds which is discuss in section 3. Section 4 describes the concept of customer relationship management (CRM). In Section 5 and 6 review the most commonly used data mining techniques in churn prediction. Finally, Section 7 concludes the paper with some future research directions.

II. CUSTOMER CHURN PREDICTION MODEL

The companies have become aware that they should put much effort not only trying to convince customers to sign contracts, but also to retain existing clients. In the current setting where people are given a huge choice of offers and different service providers to decide upon, winning new customers is a costly and hard process. Therefore, putting more effort in keeping churn low has become essential for service oriented companies. Summarize the economic value of customer retention.[15]

1. lowering the need to seek new and potentially risky customers, which allows focusing on the demands of existing customers
2. long-term customers tend to buy more
3. positive word of mouth from satisfied customers is a good way for new customers' acquisition
4. long-term customers are less costly to serve, because of a larger database of their demands
5. long-term customers are less sensitive to competitors' marketing activities
6. losing customers results in less sales and an increased need to attract new customers, which is five to six times more expensive than the money spent for retention of existing customers;
7. people tend to share more often negative than positive service experience with friends, resulting in negative image of the company among possible future customers.

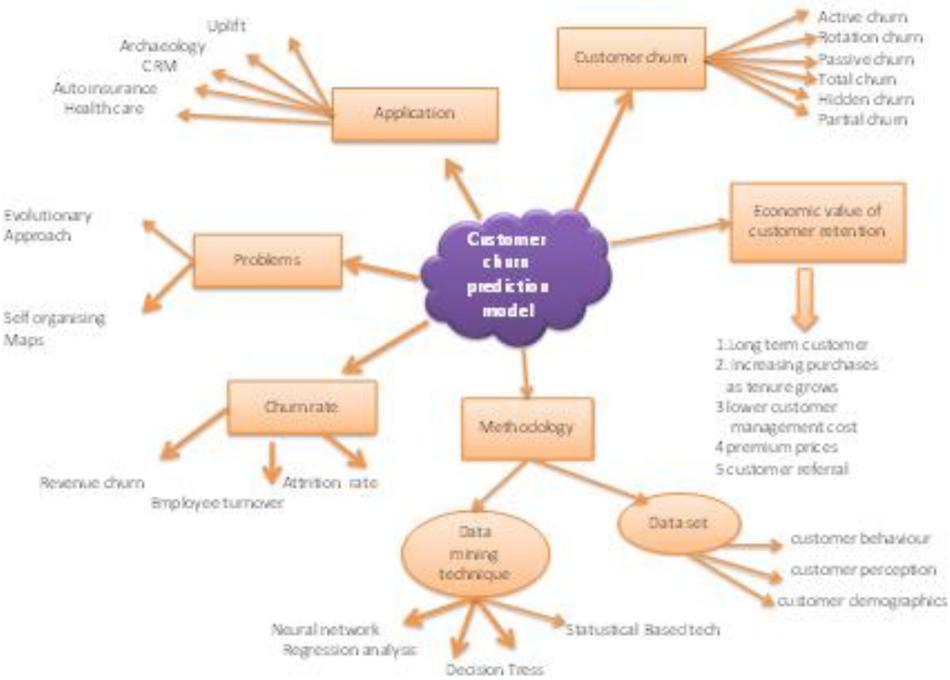


Fig 1.Customer churn prediction model

Customer Relationship Management (CRM) tools have been developed and it is applied in order improve customer acquisition and retention. Increase of profitability and to support important analytical tasks such as predictive modelling and classification. CRM applications hold a huge set of information regarding each individual customer. This information is gained from customers' activity at the company, data entered by the customer in the process of registration, calls to support hotlines,. Proper analysis of this data can bring remarkable results for the marketing purposes, but also for identifying the customers which are likely to cancel their contract. Typically, database entries are scored using a statistical model defined over various attributes, which is characterize the customers. These attributes are often called predictor variables. Higher scores reveal greater possibility of churning. Models that are being built using statistical techniques like regression analysis, classification trees and neural networks

.

2.1 Customer churn

'Churn' is a word which is derived from the change and turn. It means that the discontinuation of a contract.

As shown in fig.1 there are three types of churn:

1. Active / Deliberate - The customer decides to quit his contract and to switch to another provider. Reasons for this may include the unsatisfaction with the quality of services is too high costs, not competitive price plans, no rewards for customer loyalty and no understanding of the service scheme, bad support, no information about reasons and predicted resolution time for service problems, no continuity or fault resolution, privacy concerns, etc.[15]

2. Rotational /Incidental - The customer quits contract without the aim of switching to a competitor. Reasons for this are changes in the circumstances that prevent the customer from further requiring the service, e.g. financial problems, leading to impossibility of payment; or change of the geographical location of the customer to a place where the company is not present or the service is unavailable.
3. Passive / Non-voluntary -The company discontinues the contract itself. Voluntary churn (active, rotational) is hard to predict. And while the incidental churn only explains a small fraction of overall churn it is particularly interesting to predict and react taking appropriate action to prevent deliberate churn. In order to prevent customers voluntary contract discontinuation However, the company needs to know who are the possible churners with a low probability of error in the prediction. why this specific customer has decided to leave the company for the benefit of a competitor.

Furthermore, churning can be divided also in three other groups:
1. Total - The agreement is officially cancelled;[14]
2. Hidden - The contract is not cancelled, but the customer is not actively using the service since a long period of time;
3. Partial - The agreement is not cancelled, but the customer is not using the services to a full extent and is using only parts of it, and is instead using constantly a service of a competitor.

Depending on the company, the contract type and the business model that is being applied hidden or partial churning can lead to considerable money loss (e.g. in telecommunications: the customer only pays the monthly subscription fee, but does not place a single call) and also needs to be identified and action should be taken in order not to loose completely the customer.

Moreover, it is important to the classify which of the possible churners are of further interest for the company, e.g. which customers are likely to generate more profit (these are typically customers who generated substantial revenues and then found a better offer with a good loyalty programme at a competitor), and which customers are not interesting Because, for instance, they are identified as risky. Then the company marketing department can consider direct marketing strategies in order to retain the important customers. Although churn as a whole is an unavoidable phenomenon It can be managed and the potential losses to the business can be minimized. The timely detection of possible churners, together with the effective retention efforts support this goal.

2.2 Methodology

For finding answers to the questions who and why is likely to churn a classification of customers is needed. Churn prediction deals, therefore with the identification of customers likely to churn in the near future. The basis for this is historical data, containing information about past churners. A comparison is made between these churners and the existing customers. As likely churners are identified the customers for which the classification suggests similarity to prior churners.[12]

Data Set

Service providers can easily acquire the huge volumes of data. There are present three sets of data variables: customer behaviour, customer perceptions and customer demographics.

Customer behaviour identifies which parts of the service a customer is using and how often is he using them. Interesting are product specific ownership (which product/service is owned/on loan), total product ownership (number of products owned/on loan), interpurchase time (time between the purchase of two different articles).

Customer perceptions are defined as the way a customer understand the service. They can be measured with customer surveys and include data like overall satisfaction, quality of service, problem experience, satisfaction with problem handling, pricing, locational convenience, image/reputation of the company, customer perception of dependency to the vendor, etc.;

Customer demographics are some of the most used variables for the churn prediction. They include age, gender, level of education, social status, geographical data, etc.;

2.3Churn rate

Churn rate(sometimes called attrition rate), it is a measure of the number of individuals or items moving out of a collective group over a specific period of time. The term is used in many contexts, but is most widely applied in business with respect to a contractual customer base.

Churn rate is applied to a customer base, refers to the proportion of contractual customers or subscribers who leave a supplier during a given time period. It is a possible indicator of customer dissatisfaction, cheaper and/or better offers from the competition, more successful sales and/or marketing by the competition, or reasons having to do with the customer life cycle.

2.3.1Revenue churn: Viewing churn on a customer count basis is an indicator of customer satisfaction. Another common way to look at the churn is to measuring revenue churn. Revenue churn is the monetary amount of the recurring revenue lost in a period divided by the total revenue at the beginning of the period. Revenue churn is commonly used in Software as Service Businesses (SaaS) and business models that rely on recurring revenue models.

2.3.2Employee turnover: In some business contexts, churn rate could also refer to employee turnover within a company.

© 2014, IJARCSSE All Rights Reserved

2.3.3Employee moves/attrition rate: Churn rate can also describe the number of employees that move within certain period. For example, the annual churn rate would be the total number of moves completed in a 12-month period divided by the average number of occupants during the same 12-month period. Monthly and quarterly churn rates can also be calculated.

Formula: Attrition rate (%) = (Number of employees resigned during the month / Average number of employees during the month) x 100 where Average number of employees during the month = (Total number of employees at the start of the month + Total number of employees at the end of the month) / 2.

III. RECEIVER OPERATING SYSTEM

A receiver operating characteristics (ROC) graph is a technique for visualizing, organizing and selecting classifiers based on their performance. ROC graphs have long been used in signal detection theory to depict the trade off between hit rates and false alarm rates of classifiers. A Receiver Operating Characteristic (ROC) chart is a two-dimensional plot with the proportion of false positives (1- specificity) on the horizontal axis and the proportion of true positives on the vertical axis (sensitivity) when using different cut-off for a classifier score. Each point on the ROC curve represents a sensitivity or specificity pair corresponding to a particular decision threshold. The optimal balance point between sensitivity and specificity can be determined using the graph. A common measure for comparing the accuracy of various classifiers is the area under the ROC curve (called AUC)[1]. It evaluates the method's ability to correctly classify. The classifier with the greatest area under curve will be considered better. The closer to 1 is the AUC of a classifier and the higher accuracy.

Recent years have seen an increase in the use of ROC graphs in the machine learning community. Due in part to the realization that simple classification accuracy is often a poor metric for measuring performance. In addition to being a generally useful performance graphing method that they have properties that make them especially useful for domains with skewed class distribution and unequal classification error costs. These characteristics have become increasingly important as research continues into the areas of cost-sensitive learning and learning in the presence of unbalanced classes. Receiver operating characteristic (ROC) graphs are conceptually simple[2], but there are some non-obvious complexities that arise when they are used in research. There are also common misconceptions and pitfalls when using them in the practice.

3.1Classifier performance

By considering classification problems using only two classes. Formally, each instance I is mapped to one element of the set {p,n} of positive and negative class labels. A classification model (or classifier) is a mapping from instances to predicted classes. Some classification models produce a continuous output (e.g., an estimate of an instances class membership probability) to which different thresholds may be applied to predict class membership. Other models produce a discrete class label indicating only the predicted class of the instance. To distinguish between the actual class and the predicted class we use the labels{Y,N} for the class predictions produced by a model. Given a classifier and an instance, there are four possible outcomes. If the instance is positive and it is classified as positive, it is counted as a true positive; if it is classified as negative, it is counted as a false negative. If the instance is negative and it is classified as negative, it is counted as a true negative; if it is classified as positive, it is counted as a false positive. Given a classifier and a set of instances (the test set), a two-by-two confusion matrix (also called a contingency table) can be constructed representing the dispositions of the set of instances. This matrix forms the basis for many common metrics[7].

fp(False Positive)	Incorrectly identified
fn (False negative)	Incorrectly rejected
tp(True positive)	Correctly identified
tn(True negative)	Correctly rejected
Specificity	Specificity relates to the test's ability to exclude a condition correctly.
Sensitivity	Sensitivity relates to the test's ability to identify a condition correctly.
Precision	Positive predictive value
Recall	Sensitivity

True class

P n column table

True Positives	False Positives
False Negatives	True Negatives

Y

Hypothesized Class

N

$$fp\ rate = \frac{FP}{N} \qquad tp\ rate = \frac{TP}{P}$$

$$precision = \frac{TP}{TP+FP} \qquad recall = \frac{TP}{P}$$

$$accuracy = \frac{TP+TN}{P+N}$$

$$F-measures = \frac{2}{1/_{precision} + 1/_{recall}}$$

Fig 2. Confusion matrix

© 2014, IJARCSSE All Rights Reserved

$$tp\ r\quad ate \approx \frac{Positve\ correctly\ classified}{Total\ positives}$$

Figure shows a confusion matrix and equations of several common metrics that can be calculated from it. The numbers along the major diagonal represent the correct decisions made, and the numbers of this diagonal represent the errors—the confusion—between the various classes. The true positive rate1 (also called hit rate and recall) of a classifier is estimated as

The false positive rate (also called false alarm rate) of the classifier is

$$fp\ rate \approx \frac{Negatives\ incorrectly\ classified}{Total\ negatives}$$

Additional terms associate with ROC curves are

$$sensitivity = recall$$

$$specificity = \frac{True\ negatives}{False\ positives + True\ negatives}$$

= 1- *fp rate*

Positive predictive value= precision

IV. CUSTOMER RELATIONSHIP MANAGEMENT

4.1. Definition

Customer Relationship Management is defined by four elements of a simple framework: Know, Target, Sell, Service . CRM requires the firm to know and understand its markets and customers. This involves detailed customer intelligence in order to select the most profitable customers and identify those no longer worth targeting.

CRM also entails development of the offer: which products to sell to which customers and through which channel. In selling, firms use campaign management to increase the marketing department's effectiveness. Finally, CRM seeks to retain its customers through services such as call centres and help desks.

CRM is essentially a two-stage concept. The task of the first stage is to master the basics of building customer focus. This means moving from a product orientation to a customer orientation and defining market strategy from outside-in and not from inside-out. The focus should be on customer needs rather than product features.

Companies in the second stage are moving beyond the basics; they do not rest on their laurels but push their development of customer orientation by integrating CRM across the entire customer experience chain, by leveraging technology to achieve real-time customer management, and by constantly innovating their value proposition to customers.

4.2. Components of customer relationship management

Customer relationship management is a combination of several components. Before the process can begin, the firm must first possess customer information. Companies can learn about their customers through internal customer data or they can purchase data from outside sources. There are several sources of internal data:[12]

- summary tables that describe customers (e.g., billing records)
- customer surveys of a subset of customers who answer detailed questions
- behavioural data contained in transactions systems (web logs, credit card records, etc). .

An enterprise data warehouse is a critical component of a successful CRM strategy. Most firms have massive databases that contain marketing, HR, and financial information. However, the data required for CRM can be limited to a marketing data mart with limited feeds from other corporate systems. Experience with CRM will dictate when to aggregate data for reasons of simplicity and when to keep the data granular. External sources of data or purchased databases can be a key source for gaining customer knowledge advantage. Some examples of external data sources include lookups for current address and telephone number, household hierarchies, Fair-Isaacs credit scores, and Webpage viewing profiles.

Next, the CRM system must analyze the data using statistical tools, OLAP, and data mining. Whether the firm uses traditional statistical techniques or one of the data mining software tools, marketing professionals need to understand the customer data and business imperatives. The firm should employ data mining analysts who will be involved but will also make sure the firm does not lose sight of their original reason for doing data mining. Thus, having the right people who are trained to extract information with these tools is also important. The end result is segmentation of the market, and individual decisions are made regarding which segments are attractive.

The last component of a CRM system is campaign execution and tracking. These are the processes and systems that allow the user to develop and deliver targeted messages in a test-and-learn environment. Implementation of decisions made as a result of data mining and OLAP is done through campaign execution and tracking.

© 2014, IJARCSSE All Rights Reserved

Today there are software programs that help marketing departments handle this complex feedback procedure. Campaign management software manages and monitor customer communications across multiple touch points, such as direct mail, telemarketing, customer service, point-of-sale, e-mail, and the Web. While campaign management software may be part of the overall solution, it is primarily the people and processes that contribute to smooth interactions between marketing, information technology, and the sales channels.

V. DATA MINING AND CUSTOMER RELATIONSHIP MANAGEMENT

It should be clear from the discussion so far that customer relationship management is a broad topic with many layers, one of which is data mining, and that data mining is a method or tool that can aid companies in their quest to become more customer-oriented. Now we need to step back and see how all the pieces fit together.

5.1. The relationship

The term "customer lifecycle" refers to the stages in the relationship between a customer and a business. It is important to understand customer lifecycle because it relates directly to customer revenue and customer profitability. Marketers say there are three ways to increase a customer's value: increase their use (or purchases) of products they already have; sell them more or higher-margin products; and keep the customers for a longer period of time.

However, the customer relationship changes over time, evolving as the business and the customer learn more about each other. So why is the customer lifecycle important? Simply put, it is a framework for understanding customer behaviour. In general, there are four key stages in the customer lifecycle[13]:

1. ***Prospects***—people who are not yet customers but are in the target market
2. ***Responders***—prospects who show an interest in a product or service
3. ***Active Customers***—people who are currently using the product or service
4. ***Former Customers***—may be "bad" customers who did not pay their bills or who incurred high costs; those who are not appropriate customers because they are no longer part of the target market; or those who may have shifted their purchases to competing products[12].

The customer lifecycle provides a good framework for applying data mining to CRM. On the "input" side of data mining, the customer lifecycle tells what information is available. On the "output" side, the customer lifecycle tells what is likely to be interesting.

Looking first at the input side, there is relatively little information about prospects except what is learned through data purchased from outside sources. There are two exceptions: one, there are more prospecting data warehouses in various industries that track acquisition campaigns directed at prospects; two, click-stream information is available about prospects' behaviour on some websites. Data mining can predict the profitability of prospects as they become active customers, how long they will be active customers, and how likely they are to leave.

In addition, data mining can be used over a period of time to predict changes in details. It will not be an accurate predictor of when most lifecycle events occur. Rather, it will help the organization identify patterns in their customer data that are predictive. For example, a firm could use data mining to predict the behaviour surrounding a particular lifecycle event (e.g., retirement) and find other people in similar life stages and determine which customers are following similar behaviour patterns .

The outcome of this process is marketing data intelligence, which is defined as "Combining data driven marketing and technology to increase the knowledge and understanding of customers, products and transactional data to improve strategic decision making and tactical marketing activity, delivering the CRM challenge"[7]. There are two critical components of marketing data intelligence: customer data transformation and customer knowledge discovery. Raw data extracted and transformed from a wide array of internal and external databases, marts or warehouses and the collecting of that total data into a centralized place where it can be accessed and explored is data transformation. The process is continued through customer knowledge discovery, where the information is mined, and usable patterns and inferences can be drawn from the data. The process must be measured and tracked to ensure that the results fed to campaign management software produce information that the models created by data mining software find useful and accurate .

Data mining plays a critical role in the overall CRM process, which includes interaction with the data mart or warehouse in one direction, and interaction with campaign management software in the other direction. In the past, the link between data mining software and campaign management software was mostly manual. It required that physical copies of the scoring from data models be created and transferred to the database. This separation of data mining and campaign management software introduced considerable inefficiency and was prone to human error. Today the trend is to integrate the two components in order to gain a competitive advantage.

Firms can gain a competitive advantage by ensuring that their data mining software and campaign management software share the same definition of the customer segment in order to avoid modelling the entire database. For instance, if the ideal segment is high-income males between the ages of 25 and 35 living in the northeast, the analysis should be restricted to just those characteristics. In addition, the selected scores from the predictive model should flow directly into the campaign segment in order to form targets with the highest profit potential.

5.2. Data mining and customer privacy

While data mining techniques help businesses address more questions than ever before, this capability may add to the risk of invading customer privacy. On one hand, mining customer data can help build an intimate relationship between businesses and their customers. On the other, databases can be used against customers' wishes or to their detriment. However, personalization of CRM is far from invasion of an individual's privacy. Personal information collected by businesses can be classified in two categories: data that are provided and accessible to the users, and data that are generated and analyzed by businesses. Before data mining became popular among businesses, customers' data was generally collected on a self-provided or transactional basis. Customers themselves provide general descriptive data which contains demographic data about themselves. Transactional data refers to data obtained when a transaction takes place, such as product name, quantity, location, and time of purchase. These data are collected from registration forms, order forms, computer cookies, log files, surveys, and contests[13].

The power of data mining helps turn customer data into customer profiling information. This kind of information belongs to the second category and is accessible to businesses, although this fact may not be known to consumers. It may include customer value, customer targeting information, customer rating, and behaviour tracking. Once this information is obtained by marketers or businesses, consumers may periodically receive timely and personalized information. However, when abused, people may also suffer from certain forms of discrimination (such as insurance) or loss of career. Without proper scrutiny when applying and releasing profiling information, consumers may turn away from any effort to maintain a closer customer relationship. The central issue of privacy is to find a balance between privacy rights for consumer protection and while still providing benefits to businesses. Several advocacy groups and private efforts have been formed to promote the responsible use of technology for personalizing consumer and business relationships.

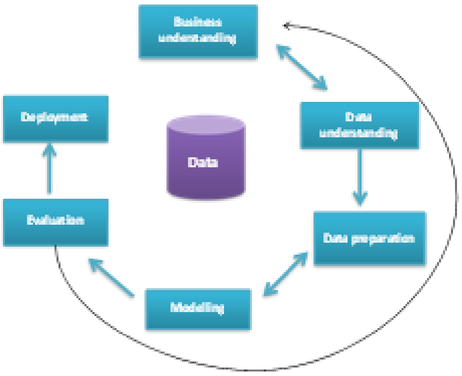
However, privacy is more of a policy issue than a technology one. One basic principle for businesses using personalized technology is to disclose to their consumers the kinds of information they are seeking and how that information will be used. Some groups list objectives for ethical information and privacy management. Others have developed a Privacy Bill of Rights that includes fair access by individuals to their personal information, responsible linkage of online and off-line information, suitable criteria for opt-in and opt-out privacy options, standardizing the disclosure to consumers of any existing privacy policy, independent verification of implementation and execution of privacy and security policies, and fair mechanisms for resolving disputes by a trusted third party.

Customer privacy can be better protected when customers do not have to reveal their identities and can remain anonymous even after data mining probing. One way to achieve this goal is to create an anonymous architecture for handling customer information. In this architecture, identity information is processed with an additional encryption procedure whenever data are fed into a data mining module for analysis. The encrypted identity information remains unique for each individual but does not diminish the power of data mining while keeping the customer's identity information protected under a firewall. Some third-party organizations also take responsibility for handling identity information, becoming a surrogate for executing targeted marketing efforts, such as mail promotion messages to the targeted individual.

VI. DATA MINING TECHNIQUES

In the last few decades there have been significant improvement and changes in the data volumes stored in files, databases, and other repositories. To aid in the decision-making process, it is necessarily vital to come up with powerful techniques of data analysis and interpretation as well as develop tools that can be important in the extraction of interesting hidden patterns and knowledge. Data mining algorithm has the capability of unveiling these patterns and their hidden relationship. It is an integral component of a complex process that is commonly known as the Knowledge Discovery in Databases which explains the steps that must be taken to ensure comprehensive data analysis[7]. CRISP-DM model stands for Cross-Industry Standard Process for data mining model. It is mainly for conducting a data mining process, whose life cycle consists of six phases as shown in Fig

The first step is to understand the data that serves commercial values. Data preparation entails preprocessing of the raw data containing limited information. This may sometimes involve removal of missing values, quantizing, conversion of categorical variables into numerical. The modeling process involves building a suitable model used to extract the information. It also evaluate the information to serve business purposes and accepting the same model after checking for important attributes like performances and accuracy[12]. The final stage involves the generation of a report or implementing a repeatable data mining process across the entire firm involved as a deployment and last phase



Application of data analysis to churn is targeted towards prediction of whether an individual customer will churn. The time that churn is expected to happen and reasons for which the churn takes place. Through prediction of the customers that are most likely to churn. The telecommunication companies are able to cut down the rate of churn through offering customers alternative and better incentives or packages to find reasons to stay. To successfully manage the churn prediction challenge and different researchers have put into use different machine-learning algorithms in addition to the data mining tools. This section presents the major data mining methods that are neural networks, statistical based techniques, decision trees, and covering algorithms and their usage in the context of customer churn analysis.

6.1 NEURAL NETWORKS

Neural Networks is the data mining technique that has the capability of learning from errors. Neural Networks are motivated by the brain. This happens in the sense that the brain learns a few new things which then will be transmitted via neurons. Equally, the neural network neuron with the learning algorithms is able to learn from training data. This makes them be referred to as Artificial Neural Networks (ANN)[10] . The results of the Lazarov and Capota[14] work showed that artificial neural networks gave the best results as compared to other known algorithms. Moreover they argued that an appropriate prediction model requires constant updating, and it should put into application a variety of data mining algorithms. Au et al. believe that largest limitation of neural networks is that they hardly uncover patterns in an easily understandable manner. Their study also had shown that neural networks outdoes decision trees for prediction of churn through identification of more churners compared to C4.5 decision trees. This is in line with the research provided by Mozer et al. in which shows that the nonlinear neural network outdoes the decision tree and logistic regression. Sharma and Panigrahi propose a neural network-based approach in the prediction of customer churn in line with cellular wireless services. The outcomes of experiments on a churn dataset of UCI repository indicate that neural network based approach can predict customer churn with accuracy more than 92%. Accuracy that is achieved by neural networks fully outweighs the limitation that they need large volumes of data sets and a lot of time to calculate a considerable load for the predictor attributes.

6.2 STATISTICAL BASED TECHNIQUES

Statistical techniques are a collection of methods applied in data mining used to process large volumes of data. They are used in learning links between both the dependent and independent attributes. This section presents the major statistical based data mining techniques (Linear regression, Logistic regression, Naive Bayes Classifier and K-nearest neighbours algorithm) and their usage in the context of customer churn Analysis.

Techniques based on regression have been associated with good results in prediction and estimation of churn. In Customer churn problem, there is often a two decisions' categorical outcome. The result is Yes or No or true or false or churns or no churns. The remaining variables are mostly continuous in nature because of that logistic regression appeared to be the best choice. [14]Lazarov & Capota discussed commonly used data mining algorithm in customer churn analysis and prediction. Regression tree techniques were discussed along with other popular data mining methods like Decision Trees, Rule based learning and Neural Networks. The conclusion was that good prediction models have to be constantly developed and a combination of the proposed techniques has to be used. Qureshi et al. also applied logistic regression techniques on telecom industry data to identify churners. It failed to perform well because only 45% of the total number of churners were correctly identified which is a very low percentage. On the contrary, the logistic regression did a good job by identifying 78% of the total number of active users correctly. Another application is done by Nie et al. who used two data mining algorithms; decision trees and logistic regression to construct a churn prediction model. They used credit card data from a real Chinese bank. The test result graded regression ahead of decision trees. Naive Bayes is a supervised learning module which makes predictions about unseen data based on Bayesian theorem . [15]Nath & Behara came up with a prediction model of customer churn. This was based on Naive Bayes algorithm in wireless customer data. It obtained 68 % accuracy in the first pass that was based on the Bayesian model. K-nearest Neighbours algorithm is one of the basic traditional statistical classification approaches. The class label assignment of the unseen instance is based on the dominant class label of the k neighbour instances. This classifier consider only the k closest entries in the training set. Zhang et al. who presented in their research a hybrid approach of the k-nearest neighbour algorithm and also the logistic regression method for constructing a binary classifier named KNN-LR. They carried out a comparison between KNN-LR with logistic regression, C4.5 and radial basis function (RBF) network. The result was that KNN-LR outperformed RBF on all the four benchmark datasets. In addition, it also outperformed logistic regression on these benchmark data sets, only that they have very close performance on the Wisconsin breast cancer data set. The outcome also indicated its superiority over RBF and C4.5 but C4.5 just exceeded KNN-LR on telecom dataset. The novel model presented by Huang & Kechadi indicates a hybrid model that joins a modified k-means clustering algorithm with a classic rule inductive technique (FOIL) for predicting customer churn behaviour. A comparison was done to the model based on six techniques. These were original k-means, decision tree, logistic regression, PART, SVM, KNN, and OneR and other Hybrid techniques like k-NN-LR, SePI. Out of all these six classifiers, hybrid models and benchmark datasets, the proposed system was 12 times better. There was then the computation of the average AUC values (measurement of prediction accuracy) for each classification technique, and the hybrid model still has the maximum average value.

6.3. DECISION TREES

Decision trees are the most common methods used in predicting and evaluating the classification of customer churn problems. Decision trees are developed using the concept of divide-and-conquer. To evaluate a customer's dataset by developing a decision tree the classification is done by altering the tree until a leaf node is attained. When evaluating a customer record a value of churner or non churner is assigned to its leaf node[12].

© 2014, IJARCSSE All Rights Reserved

There are many covering algorithms families like AQ, CN2, RIPPER, and RULES family where rules are directly induced from a given set of training examples. This can be illustrated using Verbeke et al. application of two novel data mining methods to customer churn prediction. They also benchmarked to ancient rule induction techniques for example C4.5, RIPPER, SVM, and logistic regression. They used both ALBA and AntMiner+ to stimulate accurate and understandable rules for classification. The experiments results proved that in order to get the highest accuracy a combination of ALBA with C4.5 or RIPPER is needed. If C4.5 and RIPPER are applied on an oversampled dataset the sensitivity will be on the highest level. RULE Extraction System (RULES) was distinguished from the other covering algorithms families because of its simplicity[15]. The first member of RULES family of algorithms RULES-1, has been published in 1995. After that several versions of the algorithm have been developed and applied in several domains. From the literature review, we found out that there has been little research work on inductive learning covering algorithms and their applications in customer Churn in telecom industry. RULES family algorithms are very suitable tools for data mining applications. For example, Aksoy et al. have stated that RULES-3 Inductive Learning Algorithm is a very good choice for data mining. In a study they used RULES-3 on eleven real life data sets for data mining by comparing it with three statistical, two lazy, and six rule-based data mining algorithms in terms of learning rate, accuracy and robustness to noisy and incomplete data. The good performance of RULES-3 is because of its following features: RULES-3 can handle a large sets of examples without having to break them up into smaller subsets; it can produce only rules that contain only relevant conditions; it allows a degree of control over the number of rules to be extracted; it could be applied to problems involving numerical attributes as well as nominal attributes and it gives a high flexibility for the user to control the precision of the rules to be generated, which can help in building better models.

VII. CONCLUSION

Customer churn has been identified as a major problem in all the industry and aggressive research has been conducted in this by applying various data mining technique. Different data mining techniques are generally are applied in customer churn. There are likely to be tremendous rates of research in data mining and their application in customer churn, but still it is an active research field and researchers are searching for more accurate solution. In this paper we provide a summary of the different data mining methods, and their application in customer churn prediction.

REFERENCES

[1] T. Fawcett, “An Introduction to ROC Analysis,” Pattern Recognition Letters, vol. 27, no. 8, pp. 861-874, 2006.

[2] D. Hand, “Measuring Classifier Performance: A Coherent Alternative to the Area under the ROC Curve,” Machine Learning, vol. 77, no. 1, pp. 103-123, 2009.

[3] W. Verbeke, D. Martens, C. Mues, and B. Baesens, “Building Comprehensible Customer Churn Prediction Models with Advanced Rule Induction Techniques,” Expert Systems with Applications, vol. 38, pp. 2354-2364, 2011.

[4] A.A. Nanavati, R. Singh, D. Chakraborty, K. Dasgupta, S. Mukherjea, G. Das, S. Gurumurthy, and A. Joshi, “Analyzing the Structure and Evolution of Massive Telecom Graphs,” IEEE Trans. Knowledge and Data Eng., vol. 20, no. 5, pp. 703-718, May 2008.

[5] S. Ali and K. Smith, “On Learning Algorithm Selection for Classification,” Applied Soft Computing, vol. 6, no. 2, pp. 119-138, 2006.

[6] P. Baldi, S. Brunak, Y. Chauvin, C. Andersen, and H. Nielsen, “Assessing the Accuracy of Prediction Algorithms for Classification:An Overview,” Bioinformatics, vol. 16, no. 5, pp. 412-424, 2000.

[7] N. Chawla, “Data Mining for Imbalanced Datasets: An Overview,” Data Mining and Knowledge Discovery Handbook, pp. 875-886, Springer, 2010.

[8] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, “New Insights into Churn Prediction in the Telecommunication Sector: A Profit Driven Data Mining Approach,” European J. Operational Research, vol. 218, no. 1, pp. 211-229, 2012

[9] R. Prati, G. Batista, and M. Monard, “A Survey on Graphical Methods for Classification Predictive Performance Evaluation,” IEEE Trans. Knowledge and Data Eng., vol. 23, no 11, pp. 1601-1618, Nov. 2011.

[10] Z. Zhou and X. Liu, “Training Cost-Sensitive Neural Networks with Methods Addressing the Class Imbalance Problem,” IEEE Trans. Knowledge and Data Eng., vol. 18, no. 1, pp. 63-77, Jan. 2006

[11] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, “New Insights into Churn Prediction in the Telecommunication Sector: A Profit Driven Data Mining Approach,” European J. Operational Research, vol. 218, no. 1, pp. 211-229, 2012.

[12] Amal M. Almana et al Int. Journal of Engineering Research and Applications www.ijera.com ISSN : 2248-9622, *Vol. 4, Issue 5(Version 6), May 2014, pp.165-171* www.ijera.com 165 page “A Survey On Data Mining Techniques In Customer Churn Analysis For Telecom Industry”.

[13] M. Clemente, V. Giner-Bosch, S. San Matias “Assessing classification methods for churn prediction by composite indicators”Department of Applied Statistics, Operations Research and Quality, Universitat Politecnica deValencia, Camino de Vera s/n, 46022 Spain.

[14] V. Lazarov and M. Capota, “Churn Prediction,” *TUM Comput. Sci.*, 2007.

[15] S. V Nath and R. S. Behara, “Customer churn analysis in the wireless industry: A data mining approach,” in Proceedings-Annual Meeting of the Decision Sciences Institute, 2003.

© 2014, IJARCSSE All Rights Reserved