Customer Churn Prediction using Predictive Analytics in Telecommunication Market: A Review

Yasser Khan, Shahryar Shafiq, Gohar Saleem Parvaiz, Sheeraz Ahmed, Nadeem Safwan, SafiUllah Khan
Iqra National University, Peshawar Pakistan
Institute of Management Sciences Peshawar
Pakistan Telecommunication Company Limited
Corresponding Author: yss_kh@yahoo.com

Abstracts --- In the face of extreme competitive telecommunication market, the cost of acquiring new customer is much more expensive than to retain the existing customer. Therefore, it has become imperative to pay much attention towards retaining the existing customers in order to get stabilize in market comprised of vibrant service providers. In current market, a number of prevailing statistical techniques for customer churn management are replaced by more machine learning and predictive analysis techniques. This article reviews the customer churn prediction problem, factors escalating the phenomena, prediction through predictive analytics, steps for processing of predictive analytics and evaluation of performance metrics for various churn prediction models are surveyed. Moreover, the CRM data from Pakistan Telecommunication Company limited as case study to discuss the process of data mining and predictive analytics for customer churn prediction.

Keywords- Predictive analytics, Telecom customer churn and retention, customer relationship management

I. INTRODUCTION

The public services has improved significantly on the emergence of mobile phone technology and delivered benefits to common citizen in term of connectivity, access to markets and social integration. This trends in this latest technology, however, has brought a score of demands and complaints regarding service as well [1].

In today competitive world of telecommunication technology, the surviving of operators is quite difficult due to frequent movement of customers who are discarding a particular service on basis of dissatisfaction or acquiring of offering from another service provider. Technically, this phenomena is termed as churn that has help customer obtained better service on one hand but lead significant revenue or profit loss to that network providers invested in one of expensive market. The re-acquisition of this defected customer and many more is big challenge for the company to strengthen its position in market. Therefore, companies are applying all possible efforts by applying cutting edge technology and computer application in advance to highlight those customers likely to switch over to their competitors and prevent themselves from significant loss in term of potential revenue/profit loss. So, the entire process of identification of dissatisfied customer is called churn prediction. Therefore, computerized search is necessary to highlight those customers in advance intending to end the subscription to an operator and initiate retention program [2]. By customer churn prediction, the company is assigning probability to customers on the basis of historical record and future churning behavior before it actually exhibit high propensity to switch over to competitor. Some company are also ranking their subscribers into category from high to low likely churn customers and marketing team is involved to start retention campaign to high risk customer. The behavior of potential churn customers can be modeled as most susceptible to marketing campaign and those who will churn anyway, whether they are subject to extensive marketing actions are termed as persuadable and non-persuadable customers respectively and the phenomena is known as net effect or uplift modeling [3]. The churn can be reduced by offering all possible features possible in a particular service or product and giving customer centric approach by revising all traditional strategies to face the fierce competition in tough market operating under dynamic condition with competitive offerings [4].

The identification of potential churn customers by application of predictive analytics can add substantial amount of profits compare to random selections of these customers so, wastage of resources are controlled and targeted retention
Although successful in
approach to the problem, service provision (depend upon
the existence of a particular service process are service
providers are in the business of parallelization of
networks. The model has
been developed for the prediction of churn from beginner level of players to
maturity and ultimately leave the gaming industry. This method is
applicable to score of variables besides gaming industry and help in online forecasting for parallelization of algorithms [7].
In contrast to conventional method for churn prediction
methods normally used for telephone service provider in
saturate telecom market, the model is devised and focused on
profit maximization by applying new technique termed as
expected maximum profit for customer churn. The technique is
strengthen through application of profLogit classifiers that is
similar to logistic model and genetic algorithm. The model has
attained excellent position in profit based precision for nine real
life data sets [8].The customer retention has significantly
improved both strategically and analytically despite fierce
competition that has allowed to emerge new system CRM (Customer Relationship Management). The big advantage
currently provided by CRM is classification and prediction of
potential customers about to leave the company from big data
set. Despite the fact that several models were developed for
customer retention but all have some serious limitation and
remain unable to determine the accurate and precise prediction.
The main reason behind the false result in previous model is due
to application of statistical technique (although successful in
diverse domain), validation through benchmarking of data sets
having noisy and missing values in data, true positive(TP) is
totally ignored which classified between churners and non-
churners, non-application of fuzzy logics [9].

By applying relational learner and employing social
network analytics to determine the customer churn prediction
from nine data sets obtained has significantly improved the
performance of predictive modeling. The statistical evaluation
of relational learners and model accuracy in combination with
traditional methods, network construction and application of
weights and edges in network for optimal prediction of churn in
saturated telecom industry [10].The impact service quality,
price and convenience on of customer satisfaction in one study
by taking the response from on questionnaire drawn on five
point likert scale. The customer satisfaction was modeled with
different customer segments to determine the customer
behavior. The conclusion found was that customer retention
mainly depend on socio-demographic characteristic and
customer satisfaction. Also, declared the customer behavior as
immature until the industry has not finished evolution stage
[11].The relationship of subscriber churn and customer loyalty
is determined in another study in Korean telecommunication
market by applying binomial logic model. The variables used
in study are call quality, brand image, income level, call quality
and the period the customer remain in subscript with a company
to seek the switching behavior of customers [12].In another
study the author has collected the responses from over 480
customers of telecom industry in Taiwan where the relationship
of CRM (customer relationship management) and customer
loyalty is established with controlling variable of brand image.
Also, demographic information is also collected. The research
is concluded on strong positive correlation between customer
relationship, brand image and loyalty [13].

Similarly, in pursuit of factors affecting customer churn in
telecom sector, the relationship of customer service and
customer retention is formed with mediating effect of customer
satisfaction. The data from Nigerian mobile service provider
was collected and analyzed to predict the perceived value,
customer service and satisfaction in order to predict the
customer behavior. The result clearly reveal that if customer are
provided with quality of service, they will continue their
relationship with existing network and would return if they get
defected by chance [14]. By taking other variables like price,
facility, communication, customer service and customer
satisfaction to find the customer perception to choose which
mobile service is likely to fulfill the need of the customer was
research conducted in India. The model was constructed having
three major variable customer care (depend upon complaint
management system), price (depend upon promotional offers,
call rate and product range), service provision (depend upon
call connectivity, call quality and drop rate in a specific
geographical area). The author concluded the research by
declaring service provision and then price are main factors
greatly impact customer decision and lastly the customer care
[15].The switching behavior to competitor was analyzed in
another study conducted at Chennai, India that influence
customer decision. Author has made good attempt in his article
to prove that churning decision from a particular service
provider were not influence by advertisement campaign rather
family decision. The amount of family decision to switch over
is extracted as 41% as compared to only 2% by advertisement.
Although others factors influence over all process are service
problem, customer service, service usage and cost. He also
declared call rate as important decision making variable after
network coverage [16]. The time for telemarketing is decisive
factors for customer retention especially when the contract
between service provider and customer is near to expire. The
author, however, declared varies from one group to another
group during his studies. The samples were taken in different
groups depend on monthly billing charges, hence almost 400
customers were randomly selected and research is concluded
[17]. The switching behavior was further studied taking service
quality, customer demographics, customer satisfaction and
relational investment as independent variables impacting

[98]
customer churn. The behavior was also researched with mobile network portability (MNP) which is positively affecting the churning behavior [18].

In another research analysis, almost 1000 prepaid subscribers underwent the studies from Andhra Pradesh, India, used demographic profiles, monthly billing, monthly usage, handset used and customer satisfaction. The statistical analysis conducted by author are independent sample T-Test and Discriminant analysis reveal the direct impact of call tariffs, customer care, network coverage on churning behavior of customer [19]. In order to determine the churning impact, the sixteen years historical data obtained from Newspaper Company was considered for predicting the churn. The relationship of socio demographic, frequency, fee and duration related parameters were established with churn by using classification and logistic regression [20].

III. PREDICTIVE ANALYTICS OVERVIEW

This technique is basically used to extracts hidden data patterns from big data bases so briefly know as knowledge discovery in data base. DM is primarily comprised of knowledge system derived from data base systems, machine learning, statistics and artificial intelligence and can explained as identification of interesting patterns and transformation into understandable manner in large data base.

![Fig. 1. The Predictive Modeling Techniques](image)

Firstly, the data base system are varied according to function, they performing in data mining approach like spatial, multimedia, legacy, web, relational and transaction data base. Secondly, categorization of data mining is according to techniques used for data mining like data drive, autonomous knowledge, interactive, generalized, mathematical-based, pattern based and underlying data mining approach. Thirdly DM can be categorized on the basis of knowledge. Below are most famous technique used are Association, Clustering, prediction and sequential patterns. In association, each data item is classified into predefined sets of groups, clustering makes classification according similar characteristics items, prediction is used to established connection between dependent and independent variables while sequential patterns seeks to highlight same pattern data or transaction data over specific period of time.

![Fig. 2. Different Approaches used for Data Mining](image)

<table>
<thead>
<tr>
<th>TABLE 1: METHODS OF DATA MINING</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Function</strong></td>
</tr>
<tr>
<td><em>Classification</em></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><em>Estimation</em></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><em>Segmentation</em></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><em>Association</em></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><em>Prediction</em></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

[99]
IV. PREDICTIVE ANALYTICS TECHNIQUES

A. Decision Tree

For classification of customer churn problem, this method is most commonly used. The customer data set is classified by continuously dividing the tree until reach at leaf node. Hence the rule of divide and conquer is used in the development of decision tree is used so by evaluation the record of customers obtained from a specific telecom operator when reach to leaf node would complete the process of decision tree. The technique although has some limitation in addressing complex and non-linear connection in attributes. In order to enhance the performance of the algorithm, the method is frequently used in combination with other techniques like with neural network.

The minor causes for customer churn is to be ignored while the area of subscribers is the main feature for classification hence application of decision tree for classification is used to determine the customer churn analysis in telecommunication industry [21].

B. Logistic Regression

The conditional or posterior probability of discrete value is computed for generation of continuous or discrete variables. The standard error rate is calculated for each variable in this modeling. On the basis of significance to linear regression for churn prediction LG model is designed and constructed. The main aim behind logistic regression is identify either the customer is going to be churner or non-churner. Logistic regression

\[
\pi_i = \frac{1}{1 + e^{-(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_k x_k)}}
\]

Where \(\pi_i\) is the outcome probability, \(\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_k\) are the co-efficient (Transformed into odd rations with importance degree).

C. Support Vector Machine

This technique is design mainly to reduce structural risk minimization, are developed by Boser, Guyon and Vapnik. This is supervised learning mechanism which recognized and analyzed the data set patterns for classification and regression problems.

The performance of the SVM is improved through introduction of Kernel function and a lot of emphasis is laid to develop best kernels or combination of kernels. For best data transformation, the SVM sometime surpass both neural network and Decision Tree in performance.

D. Nearest Neighbor Algorithm

The technique mostly used in data mining algorithm for predictive analytics in classification problems, estimation and prediction. In K-Nearest Neighbor, J determine the number of variables used in neighborhood. The two famous techniques used are distance function and cardinality.

Distance function

\[
y_i = \frac{1}{k} \sum_{j \in N(x_i)} y_j
\]

The neighborhood of x corresponds to xi, N (xi) and k is the constant of proportionality

In Cardinality “k” determine the complexity and importance of nearest neighbor function, whenever the k is higher in number the model may be considered as less adaptive and another circumstances k- value is used for goodness of fit test.

E. Discriminant Analysis

The technique most commonly used for linear modelling in pattern recognition and machine learning technique for convergence and divergence of objects categories. Discriminant analysis is also frequency used to find the variables that separate two or more than two variables. The DA (Discriminate...
analysis) function used to separate the variables into two groups are as below:

\[ p = \frac{1}{1 + (e^{\alpha + \beta x})^{-1}} \]

Where \( \beta = \sum (\mu_1 - \mu_0) \)

\[ \alpha = -\log \frac{\pi_1}{\pi_0} + \frac{1}{2} (\mu_1 + \mu_0) \sum (\mu_1 - \mu_0) \]

Where \( \pi_0 \) and \( \pi_1 \) are called prior probabilities, \( \mu_0 \) and \( \mu_1 \) are the distribution mean

F. Fuzzy Base Learning Algorithm

This is the most flexible, assist in modeling of non-linear function and tolerant to data which may be imprecise. This can help in implementation of conventional control techniques.

G. Neural Network

For the resolution of complex and non-linear problem, this technique has gained very much popularity. This can be both represented in hardware- based such as neuron are represented as physical components or as computer model in software form. The technique due to extensive structure has been using variety of learning algorithm and topologies. In supervised learning, when more than one hidden layers are used then model is defined as multilayer perceptron can train a scores of back propagation algorithm (BPN). Although complex shape but artificial neural network outperformed both logistic regression and C5.0 for customer churn prediction.

V. CHURN PREDICTION USING PREDICTIVE ANALYSIS MODELING

The conceptual infrastructure from conception to ultimate testing is drawing in figure:

i. The data is collected from interest group according to the behavior attributes that clearly distinguish churners from non-churners

ii. The variables are extracted, transform from given data set items

iii. The data are then analyzed for the functionality for predefined technique for our specific application. Most famous are association, clustering or prediction

iv. The model is build according to statistical or machine learning technique that is set theory, Bayesian classification, neural network, logistic regression, decision tree, genetic algorithm, linear regression etc.,

v. The already developed model is passed through testing and evaluation phase to check the accuracy.

VI. PERFORMANCE METRICS

Evaluation of a particular model is performed through performance metric which determine how effective is our classifier for a certain churn prediction. The metrics are explained as below:

A. Confusion Metrics
Model accuracy and correctness can be found through confusion metrics used frequently for classification problem. The matrix is comprised of two dimensional table (predicted and actual) along with set of classes. The actual class is represented in columns while predicted class is rows.

**TABLE 2: Confusion Matrix**

The confusion matrix will result true positive (TP) when both predicted and actual class give true values, true negative (TN) when both predicted and actual class give false values, false positive (FP) when actual class is false and predicted is true and false negative (FN) is status of CM when actual class is true and predicted is false.

B. Accuracy

This is mostly used in classification problem for correct prediction of over all kind of prediction. Mathematically, it is represented as below:

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}
\]

C. Precision

The measurement to explain the proportion of values having problem in reality. Mathematically, it is represented as below:

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

D. Recall

The measurement of proportion of real cases actually predicted by model. Mathematically, it is represented as below:

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

E. Specificity

The proportion of true results predicted as true by a particular model

\[
\text{Specificity} = \frac{TN}{TN + FP}
\]

(The exact opposite of recall is specificity)

F. F1-Score

The score a model obtained on the combination of both precession and recall. Mathematically, it is represented as below:

\[
\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

G. Receiver operating Characteristic Curve(ROC)

The curve represent true positive rate and false positive rate on x-axis and y-axis to determine both benefits and costs respectively on 1 learner scale. After having compromise between benefits and cost, this curve represent the connection churners ration identified correctly as churner while non-churner ration mistakenly recorded as churner.

Best Performance out of model is achieved when the cure passes thorough or near to 0 or 1 and sensitivity and specificity would reach to 100%. In some instances even in logistic regression , instead binary value other discreate values are derived.

![ROC Curve](image-url)
The total area of the is divided by Random line, so space above the line are considered as better performance area and on the line are random guessing classifiers. The area below the ROC curve are called area under the curve (AUC) ranges from 0.0 to 1.0 so AUC value greater than 0.5 will give out better performance area.

**H. Lift Curve**

![Lift Curve Image](image)

Fig. 8. The Lift Curve

This curve is used in the situation when true positive (TP) can not be determined due to non-availability of classification instances where ROC curve is unable to perform its due role to determine the accuracy of models. The prediction performance measuring scale used in lift curve is deciles. The first two deciles predict the information of 50% churners and first five deciles provide data of about 90% churner. The main reason behind the popularity of this metric is relative probability of churner and target for churners become more narrow when top two deciles are present for retention.

**VII. PROPOSITION OF MODEL FOR CHURN PREDICTION**

![Data Set Distribution Image](image)

Fig. 9. The Distribution of Data Set

**TABLE 3: Details of Data set Attributes**

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Description of Attributes</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>age into five groups</td>
<td>Nominal</td>
</tr>
<tr>
<td>gen</td>
<td>gender</td>
<td>Ordinal</td>
</tr>
<tr>
<td>mar_st</td>
<td>marital status</td>
<td>Ordinal</td>
</tr>
<tr>
<td>av_in_sm</td>
<td>avg. incoming call from same company</td>
<td>Scale</td>
</tr>
<tr>
<td>av_ou_sm</td>
<td>avg. outgoing call from same company</td>
<td>Scale</td>
</tr>
<tr>
<td>av_al_sm</td>
<td>avg.all call from same company</td>
<td>Scale</td>
</tr>
<tr>
<td>av_in_an</td>
<td>avg.incoming call from another company</td>
<td>Scale</td>
</tr>
<tr>
<td>av_ou_an</td>
<td>avg.outgoing call from another company</td>
<td>Scale</td>
</tr>
<tr>
<td>av_al_an</td>
<td>avg.all call from another company</td>
<td>Scale</td>
</tr>
<tr>
<td>av_in_s_a</td>
<td>avg. incoming call either of same or another company</td>
<td>Scale</td>
</tr>
<tr>
<td>av_ou_s_a</td>
<td>avg.outgoing call from either same or another company</td>
<td>Scale</td>
</tr>
<tr>
<td>av_al_s_a</td>
<td>avg. all call either of same or another company</td>
<td>Scale</td>
</tr>
<tr>
<td>av_i_o_s_a</td>
<td>avg.of incoming&amp;outgoing call either from same or another company</td>
<td>Scale</td>
</tr>
<tr>
<td>ch_mnth</td>
<td>change in use from one month to another month during observation period</td>
<td>Scale</td>
</tr>
<tr>
<td>av_sms</td>
<td>avg. number of short messaging service during observation period</td>
<td>Scale</td>
</tr>
<tr>
<td>av_rev</td>
<td>avg. monthly revenue</td>
<td>Scale</td>
</tr>
<tr>
<td>vo_&amp;_da</td>
<td>both voice and data service</td>
<td>Scale</td>
</tr>
<tr>
<td>vo_&amp;_iptv</td>
<td>both voice and multimedia service</td>
<td>Scale</td>
</tr>
<tr>
<td>vo_da_iptv</td>
<td>Triple play service</td>
<td>Scale</td>
</tr>
<tr>
<td>av_cuscar</td>
<td>avg.calls generated from customer service representative</td>
<td>Scale</td>
</tr>
<tr>
<td>av_comp</td>
<td>avg. complaints registered</td>
<td>Scale</td>
</tr>
<tr>
<td>ch_pack</td>
<td>changes in packages</td>
<td>Scale</td>
</tr>
<tr>
<td>per_bw_cal</td>
<td>time period between two call</td>
<td>Scale</td>
</tr>
<tr>
<td>churn</td>
<td>churn (yes/no)</td>
<td>Nominal</td>
</tr>
</tbody>
</table>
For analysis of data dataming tool (open source) tool called WEKA was used. Pakistan Telecommunication company limited is national internet service provider and backbone for telecom infrastructure despite a number of telecom corporations in entire country market. These services are GSM, LTE, CDMA, HSPA+, IPTV. Total 26% share of the company has been sold to Etisalate Telecommunications while remain 74% is still owned by state.

Data of almost 10,000 instances have been collected from PTCL which is divided as training and testing data as 8000 and 2000 respectively. Similarly 6400(80%) instances are labeled as non-churner and 1600 cases labeled as churner.

Out of total 2000 instances from testing data set, 1600 cases are considered as non-churner while the remaining 400 cases are taken as churners. The distribution of entire data set is illustrated in Figure-9 for pictorial view.

The given data is divided into 24 attributes according to available file taken from company for analysis illustrated in Table-4. The first section is covered by demographic information about the customers then voice service in minutes through telephonic call from same network or another network or both. This data is then feed into WEKA software and results are drawn and mentioned in details in Table-5.

A. Results Interpretation

The result interpretation of classification method and role of churn and non-churner in Decision Tree, Neural Network and Support vector machine is given in Table-5:

B. Accuracy and Error rate Comparison

By comparing the accuracy in different methodologies used for prediction of customer churn, it is declared that the best result can be extracted from neural network getting 85% accuracy and 15% error rate. Support vector machine getting 84% accuracy and 16% error rate on second position and decision tree at third position getting 78% accuracy and 22% error rate are drawn in line chart as below:

![Comparison of Accuracy and Error rate](image)

Fig. 10. The Comparison Of Accuracy And Error Rate

### Table 4: Interpretation of Results

<table>
<thead>
<tr>
<th>Confusion Matrix</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Churners</td>
</tr>
<tr>
<td><strong>Decision Tree</strong></td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td></td>
</tr>
<tr>
<td>Churners</td>
<td>810</td>
</tr>
<tr>
<td>No-Churners</td>
<td>190</td>
</tr>
<tr>
<td><strong>Neural Network</strong></td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td></td>
</tr>
<tr>
<td>Churners</td>
<td>860</td>
</tr>
<tr>
<td>No-Churners</td>
<td>140</td>
</tr>
<tr>
<td><strong>Support Vector Machine</strong></td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td></td>
</tr>
<tr>
<td>Churners</td>
<td>850</td>
</tr>
<tr>
<td>No-Churners</td>
<td>150</td>
</tr>
</tbody>
</table>

### VIII. CONCLUSION

In this paper, all related literature yet available regarding predictive analytics on customer churn is discussed in detail and explained the advantages and disadvantages of each method is mentioned. The sole purpose of this paper is not design any new algorithm for churn prediction but to accumulate and categorized the existing technique already prevailed in the competitive telecom market. The methods along with pictorial representation is furnished with required details, the entire process of data mining from data collection to model evaluation is explained in form of conceptual framework. The metrics to determine the performance of each model is mentioned in details, from confusion metrics, accuracy, precision, recall, specificity, receiver operating characteristic curve and lift curve. A case study on 10,000 randomly selected customers of Pakistan Telecom Company Limited, one of renown national corporation in Pakistan telecom market, is presented for
explanation of predictive analytics process. After applying data mining most famous techniques on data set provided and conclude that neural network is method having more accuracy and lower error rate as compare to support vector machine and decision tree. Moreover, the process of churn prediction is explained and cleared from literature review that churn prediction is pre-requisite to initiate any retention campaign and imperative for all sort of service.

REFERENCES


