Poster text

Introduction:
One of the key aspects of a successful business model is how to establish and maintain a loyal customer base. If a company is able to keep their customers in a subscription based revenue system, they are able to maintain a strong financial foundation and increase their return on investments. The cost savings can then be used to increase product technology and to create a more diverse marketing strategy to provide support to decrease customer churn. Through a stable customer base the company is able to establish a strong customer review system and increase the likelihood of customer referrals as a source of increased product subscriptions. Some of the more common issues that can increase customer churn are constantly changing product technology, the low cost for customers to switch companies, a large selection of competitors in a saturated market, and decreased oversight of corporations. The company can directly contribute to increased customer churn by providing underperforming products, increasing costs passed onto the customer, providing poor customer service and not taking advantage of targeting marketing to existing customers. If a company is able to take advantage of the large amount of data provided by the customers through customer service interactions, survey and online reviews and the base demographical data and usage they can begin to analyze where the deficiencies are and predict what customers are at more risk for attrition.

Why is this topic data science:
Being able to provide a model to help companies predict a customer's likelihood to change companies, will require analyzing and creating usable data from a large customer data set. The company will need to pull data related to demographics, contractual agreements, online reviews and product usage. Once this data is compiled it will need to have machine learning algorithms applied to look for commonalities. Using the core CRISP-DM workflow the data will need to flow through several steps to evolve from a large data set to usable data. The first step is to outline the problem and brainstorm for possible causes that are consistently associated with customer churn. Once the data is pulled from the appropriate sources, whether that be social media, online reviews, customer satisfaction surveys, product usage and base demographical data it will need to be reviewed to start exploring any discernable trends within the data. This process is typically completed through data visualization methods for histograms, pie charts and plot correlation matrix. These data visualization models will help identify trends within the data and provide the groundwork for the next step, data preparation. Data will need to be split into a

e

ze

ws

w

.

vill

d

training data set, validation data set and a test data set. The data will be normalized to help identify any outliers or redundant values that would skew the data analysis and the data will need to be balanced so the sampling is consistent with the larger data set. The next process wil be to apply modeling to the data sets with logistic regression and predict values using the previously defined value outcomes. Once the data is ready, it will be evaluated using a confusic matrix and receive accuracy score, F1 score, Precision score and a Recall score. Based on how the scores are returned (~50% or greater) the user will know if the data set is reliable to use an the predictive models are working correctly. The last part of evaluation is to review the data to see what columns are not providing value and remove any unneeded data. This modeling and predictive method are then applied to the trimmed data set to ensure it provides the same accuracy in prediction as the larger data set. Once the data is trimmed and is providing an acceptable level of predictive accuracy it can be evaluated for deployment. The scripting will need to be evaluated for its impact on the system regarding resource usage and the timeliness in pulling predictive values on the larger data sets.

What is the deliverable:

Through the use of large data sets a predictive model can be created that will evaluate trends from previously churned clients that evaluates customer service interactions, product usage, online reviews and base demographical data to provide a way for the current customer base to be analyzed and outline which clients align with the trends for churned customers. Once the company is able to identify those customers with a higher likelihood to switch companies, a targeted marketing campaign and direct customer engagement can be deployed to help reduce the overall attrition rate for the current customer base. The predictive modeling can be used fc a predetermined amount of time (i.e 3-6 months) to determine if the model is identifying the appropriate customers, and if the marketing and customer engagement methods are helping to reduce the overall customer attrition rate.

Conclusion:

Additional references
http://didawikinf.di.unipi.it/lib/exe/fetch.php/dm/churn-mobilephone.pdf

---

Predicting Customer Churn in Mobile Telephony Industry
Using Probabilistic Classifiers in Data Mining

Additional references

http://didawikinf.di.unipi.it/lib/exe/fetch.php/dm/churn-mobilephone.pdf

Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining

Clement Kirui[1], Li Hong[2], Wilson Cheruiyot[3] and Hillary Kirui[4]

[1] School of Information and Communication Engineering, Central South University
Changsha, 410083, China

[2] School of Information and Communication Engineering, Central South University
Changsha, 410083, China

[3] Institute of Computer Science and Information Technology, Jomo Kenyatta University of Agriculture and Technology
Juja, Nairobi, Kenya

[4] School of Human Resource and Development, Jomo Kenyatta University of Agriculture and Technology
Juja, Nairobi, Kenya

https://www.researchgate.net/profile/Dymitr_Ruta/publication/283992554_Churn_Prediction_Does_Technology_Matter/links/5bd1622145851537f5990142/Churn-Prediction-Does-Technology-Matter.pdf

Churn Prediction: Does Technology Matter?

John Hadden, Ashutosh Tiwari, Rajkumar Roy, and Dymitr Ruta

Abstract—The aim of this paper is to identify the most suitable model for churn prediction based on three different techniques. The paper identifies the variables that affect churn in reverence of customer complaints data and provides a comparative analysis of neural networks, regression trees and regression in their capabilities of predicting customer churn.

Keywords—Churn, Decision Trees, Neural Networks, Regression.

I. INTRODUCTION

IT has become common knowledge within companies, that their most valuable assets are their existing customers. As markets have become increasingly saturated, companies have attempted to identify ways in which to improve customer loyalty, satisfaction, and retention. The marketing approach have identified several reasons to why this type of data is unsuitable. They point out that using demographical data for the basis of churn prediction creates a churn analysis that is dependent on the customer rather than the contract. It is also suggested that demographic data held by some companies is very limited, restricting the suitability of many existing churn-prediction systems.

In response to these problems Wei and Chiu, [2] based their churn prediction model on call pattern changes and contractual information. As an alternative to this approach the authors have investigated an approach of the suitability of customer complaints and repairs data for churn prediction. The best variables are identified and neural networks, classification trees and regression are compared for their suitability for churn prediction using this type of data.

tandfonline.com/doi/abs/10.1080/10919390902821291

tandfonline.com/doi/abs/10.1080/10919390902821291

Enter keywords, authors, DOI, ORC

Original Articles

Hybrid Models Using Unsupervised Clustering for Prediction of Customer Churn

Indranil Bose & Xi Chen ✉
Pages 133-151 | Published online: 15 Apr 2009

66 Download citation   🔗 https://doi.org/10.1080/10919390902821291

📄 Full Article   🖼 Figures & data   📰 References   66 Citations   📊 Metrics   🖨 Reprints & Permissions   **Get access**

Abstract

Churn management is one of the key issues handled by mobile telecommunication operators. Data mining techniques can help in the prediction of churn behavior of customers. Various supervised learning techniques have been used to study customer churn. However, research on the use of unsupervised learning techniques for prediction of churn is limited. In this article, we use two-stage hybrid models consisting of unsupervised clustering techniques and decision trees with boosting on two different data sets and evaluate the models in terms of top decile lift. We examine two different approaches for hybridization of the models for utilizing the results of clustering based on various attributes related to services usage and revenue contribution of customers. The results indicate that the use of clustering led to

https://www.sciencedirect.com/science/article/abs/pii/S0305054805003503

📄 Get Access   Share   Export

Computers & Operations Research
Volume 34, Issue 10, October 2007, Pages 2902-2917

ELSEVIER

Computer assisted customer churn management: State-of-the-art and future trends

John Hadden [a] ✉, Ashutosh Tiwari [a] ✉, Rajkumar Roy [a] ✉, Dymitr Ruta [b] ✉

Show more ⌄

https://doi.org/10.1016/j.cor.2005.11.007   Get rights and content

Abstract

A business incurs much higher charges when attempting to win new customers than to retain existing ones. As a result, much research has been invested into new ways of identifying those customers who have a high risk of churning. However, customer retention efforts have also been costing organisations large amounts of resource. In response to these issues, the next generation of churn management should focus on accuracy. A variety of churn management techniques have been developed as a response to the above requirements. The focus of this paper is to review some of the most popular technologies that have been identified in the literature for the development of a customer churn management platform. The advantages and disadvantages of the identified technologies are discussed, and a discussion on the future research directions is offered.

https://www.sciencedirect.com/science/article/abs/pii/S1569190X15000386

A comparison of machine learning techniques for customer churn prediction

T. Vafeiadis [a] ✉, K.I. Diamantaras [b] ✉, G. Sarigiannidis [a] ✉, K.Ch. Chatzisavvas [a] ✉

Show more ∨

Get rights and content

Abstract

We present a comparative study on the most popular machine learning methods applied to the challenging problem of customer churning prediction in the telecommunications industry. In the first phase of our experiments, all models were applied and evaluated using cross-validation on a popular, public domain dataset. In the second phase, the performance improvement offered by boosting was studied. In order to determine the most efficient parameter combinations we performed a series of Monte Carlo simulations for each method and for a wide range of parameters. Our results demonstrate clear superiority of the boosted versions of the models against the plain (non-boosted) versions. The best overall classifier was the

SVM-POLY using *AdaBoost* with accuracy of almost 97% and *F*-measure over 84%.

https://link.springer.com/chapter/10.1007/978-3-642-35603-2_47

Data Mining for Churn Prediction: Multiple Regressions Approach

Authors   Authors and affiliations

Mohd Khalid Awang, Mohd Nordin Abdul Rahman, Mohammad Ridwan Ismail

Abstract

The rapid development in the telecommunications industry contributed to the increased rivalry among the competitors. Customers switch to competitors or move out from the service provider become critical concerns for companies to retain customer loyalty. Churn prevention through churn prediction is one of the methods to ensure customer loyalty with the service provider. Detect and analyze early churn is a proactive step to ensure that existing customers did not move out or subscribe to the product from competitors. Selection of customer characteristics is one of the core issues to forecast customer churn in the telecommunications industry. This paper proposes multiple regressions analysis to predict the customers churn in the telecommunications industry based on recommended features. The results have shown that the performance of multiple regressions for predicting customer churn is acceptably good.