Conquer Customer Churn with AI

Abstract

Retaining existing customers is more cost-effective than acquiring new clients. It is essential for companies to prevent departures as effectively as possible. The churn prediction model can estimate the propensity that clients will leave. By knowing which clients are at the highest risk of leaving, companies can better target rescue efforts, maintain customers with lower costs, and increase their profit.

To facilitate companies' ability to recognize possible, we use logistic regression to predict whether a customer will leave the membership based on his/her characteristics. The accuracy of the model reaches 0.81. Our work sets up a foundation for further research based on the specific need of each company in different industries.

Introduction

Customer churn is a major problem for most companies. Losing customers requires gaining new customers to replace them. This could be around 10 times more expensive than retaining existing customers, depending on the domain.

There are many ways to address churn. For some, it's a matter of having a dedicated team of customer advocacy or success experts to ensure that every customer is getting the most out of your software. Others need to continue to innovate and improve the product so customers don't walk. Others should offer incentives for customers to stay. Each approach, however, has one need in common: the ability to better understand which customers are at risk.

Finding what factor is associated with churn can help you in predicting what customers could leave next. But that's easier said than done. There may be a number of different data elements contributing to the situation — much more than human brains can absorb and correlate. This means crunching the numbers to get actionable insights will take a great deal of time and effort, not to mention the attention of a trained data scientist, depending on the size and complexity of your datasets.

That's why AI can be a difference-maker. Applying AI to the customer churn problem can enable companies to quickly identify commonalities in customer data — and apply those findings to current customers, giving customer service and sales representatives an opportunity to stop a customer from leaving before it's too late.

7 steps of Model Building

Step 1: Explain the value proposition of building the AI system for the given problem.

Customer churn is a major problem for most companies. Losing customers requires gaining new customers to replace them. This could be around 10X more expensive than retaining existing customers, depending on the domain. As a result, if the company could predict whether customers will leave based on their features, corresponding efforts could be taken to keep them subscribing to the service. By building the AI system to realize the churn prediction function, companies could benefit from maintaining their customers more efficiently.

Step 2: Explain the data set. Describe your training and test data set.

The data set is from a company's membership system. It contains 10,000 records of members' information, like surname, credit score, nationality, gender, and number of products bought. The key variable we are interested in is whether the member exited the membership ("Exited").

We randomly separate the data set into a training data set and a test data set. Specifically, we assign 80% of the data (8,000 records) as the training data and use the rest 20% as the test data.

Step 3: Explain the labels available for the given data or if you are working on unsupervised algorithm you don't need to explain Step 3.

The label, "Exited", is already presented in the data set as one dimension of each record. If the member quitted the membership, "Exited" equals one; otherwise, "Exited" equals zero.

Step 4: Explain features you extracted. If you are using structured data sets with all features are the columns of data already available, explain the motivation on using certain features.

We have 13 columns in total except the label column. Here are the features used:

Score	Credit score of the members.
Nationality	Nationality of the members. There are 3 categories: France, Spain, and Germany. We did some modifications and transferred "Nationality" to 2 binary variables: "France" and "Spain" ("Germany" is omitted as the base value). After the transformation, we deleted "Nationality" in the data set.
Gender	Gender of the members. There are 2 categories: Female and Male. We transferred "Gender" to a binary variable, "Male" ("Female" is omitted as the base value), and deleted "Gender" in the data set.
Age	Age of the members.
Tenure	Tenure of the members.
Products	Product count bought by members.
Card	Credit card ownership count of members.
Active	Active user, or not. (0/1)
Salary	Annual income of the users.

We use these features because they are logically possible to have relations with a member's decision on whether quit the membership. People with higher credit scores might less likely to quit a membership; people from different countries who are exposed to different cultures may perform different levels of consumer loyalty; males might not bother to quit; older people may be stickier to the membership; longer tenure indicates higher satisfaction and loyalty; the more products a member buys, the less likely that he/she will quit; more credit cards may indicate affection for membership holding; active members are less likely to quit; and high-income people are more capable of maintaining the membership.

We abandon “Row”, “Id”, and “Surname” in the regression because it does not make sense that they are related to “Exited”. We also do not include “Balance”, which refers to the total bank balance of the members, since there are too many empty values and members may not honestly report their bank account balances.

Step 5: Explain what algorithm you chose for training and why.

We choose Logistic Regression as the model for the following reasons:

First, since our problem is a binary classification, logistic regression is a popular algorithm in dealing with this kind of problem.

Second, it can predict probabilities and make classification based on probabilities.

Third, logistic regression is easier to implement, interpret, and very efficient to train.

Step 6: Describe the training process, parameters used, training time and accuracy of the model.

In our setting, $\{(x_i, y_i)\}_{i=1,\dots,n}$ are observed samples in the training data, where x_i is a vector of i-th client’s features, and y_i is the outcome “Exited”. We assume that

$$P(y = 1|x, \beta^*) = \frac{1}{1 + e^{-x^T\beta^*}}$$

$$P(y = 0|x, \beta^*) = \frac{e^{-x^T\beta^*}}{1 + e^{-x^T\beta^*}}$$

and minimize the negative log likelihood over the training data

$$\hat{\beta} = min_\beta \sum_{i=1}^{n} \log\left[(P(y = y_i|x_i, \beta)\right]$$

to get a $\hat{\beta}$ as the estimator of β^*. Then we plug $\hat{\beta}$ into the probability function of the test data to estimate the probability of exit. If this estimated probability is greater than a prespecified threshold value (for example 50%), our model predicts that the client will quit, and vice versa.

The training time is around 0.03s, fluctuating each time.

The accuracy of the model on the uploaded training dataset is 0.81.

	precision	recall	f1-score	support
0	0.82	0.96	0.89	6358
1	0.58	0.21	0.30	1642
accuracy			0.81	8000
macro avg	0.70	0.58	0.60	8000
weighted avg	0.77	0.81	0.77	8000

Step 7: Use the trained model for prediction. Report test accuracy.

The test accuracy is 0.82 based on the uploaded test dataset.

	precision	recall	f1-score	support
0	0.84	0.97	0.90	1605
1	0.65	0.25	0.36	395
accuracy			0.82	2000
macro avg	0.75	0.61	0.63	2000
weighted avg	0.80	0.82	0.79	2000

Solution

Our model can be viewed as the foundation of a solution. The test accuracy of our model is 0.82, which is fairly high. However, we have noticed the low recall for those who will quit. Therefore, whether this model can be applied to the real world depends on each company's requirement, in other words, the preference for the tradeoff among recall, precision, and accuracy. For future more accurate prediction and customized marketing strategies, we can collect data in more dimensions to better describe a customer. Currently, we only have 13 dimensions.

We can also change the threshold according to the industry and based on cost-benefit analysis. In our model, if we set threshold value to be 30% instead of 50%, the recall will be over 0.5 while the accuracy will be still near 0.8. Such relatively low threshold can be applied if the company need to forecast as more potential leaving clients as possible.

```
Y_pred2=2*(yhat2>0.3)/2
print(classification_report(Y_test, Y_pred2))
```

	precision	recall	f1-score	support
0	0.88	0.85	0.87	1605
1	0.47	0.53	0.50	395
accuracy			0.79	2000
macro avg	0.67	0.69	0.68	2000
weighted avg	0.80	0.79	0.79	2000

Application

With our model, the company can predict whether a customer is going to churn or not. If they are going to leave the company, then certain interventions could be conducted to get the client back, either by offering lower prices or providing more need-based products. In this way, the company can maintain customers with at a lower cost.

Challenge

There are three major challenges companies will face when implementing the AI system in the real world.

1. **Lack of a 'silver bullet' methodology**

One of the major challenges that companies face in building a predictive churn model revolves around the selection of a suitable churn modeling approach. But there is no single methodology to build a predictive churn model that can work in most situations. Machine learning techniques are mostly used by businesses due to their efficiency and ability to categorize and manipulate complex data sets. The approach of survival analysis, on the other hand, uses survival and hazard functions to predict which customer will churn during a particular period. So, the best solution to deal with this challenge is to compare the performance of several models and identify the most effective method for your business.

2. **Features and exploratory analysis**

Businesses face several roadblocks and churn risks at this stage of building predictive churn models. For example, the lack of information, target leakage, and the need for optimal feature transformations. Along with domain knowledge, businesses must also have the required skills and creativity to build robust predictive churn models. Therefore, it is important that companies execute careful exploratory analysis and build auxiliary models before embarking on building an overall churn prediction model. Exploratory analysis can also help in revealing reciprocity, irregularities, outliers, and, relationships between different functions, which wouldn't be possible with domain knowledge alone.

3. **Validating churn model performance**

For accurate customer churn analysis, it's essential to choose the correct metric to optimize and validate datasets. The precision of a churn model not only impacts performance but also affects decision-making. As such, businesses need to employ different strategies to validate the performance of a churn model prior to its implementation. Also, businesses need to monitor several versions of the churn model to identify problems.

Conclusion

Company typically focuses on acquiring new clients, then grows by offering additional products to existing clients or trying to get them to use their products more. If all is going well, there comes a point when the company is large enough that it must also choose a slightly more defensive strategy and focus on retaining existing customers. Despite the best user experience, there will always be a group of clients who are not satisfied and decide to leave. The company then faces the problem of how to prevent these (voluntary) departures as effectively as possible. This is where the churn model, among others, comes to the rescue.

Our model is a predictive model that estimates — at the level of individual customers — the propensity (or susceptibility) they have to leave. For each customer at any given time, it tells us whether the customer is higher than a certain probability (the threshold of the model; in our case, 50%) to leave and how high the risk is of losing them in the future.

Technically, it's a binary classifier that divides clients into two groups (classes) — those who leave and those who don't. In addition to assigning them to one of the two groups, it will typically give us the probability with which the client belongs to that group. It is important to note that this is the probability of belonging to the group of clients who leave. Thus, it is the propensity to leave and not the probability of leaving. However, it is possible to estimate the probability through a churn model.

By knowing which clients are at the highest risk of leaving, we can better target our rescue efforts. For example, we can reach out to these clients with a marketing campaign, reminding them that they haven't purchased from us in a while, or even offering them a benefit.