Case Paper: Predicting Customer Churn at QWE Inc.

Austin Koenig

1 Background

QWE Inc. offers a subscription service to small and medium-sized companies to manage their online branding either on a month-by-month or 6-/12-month basis. Richard Wall, VP of Customer Services at QWE, asked V.J. Aggrawal to take on a project that utilizes available data in order to change the way in which QWE handles customer churn. Wall's hope was to move from their previous reactive approach towards a more proactive one.

The old process of handling churn started when the customer called to cancel their subscription, at which point the customer service representative would go on enticing them to prolong their contract using incentives like free services or discounts on existing ones. This paper is not considering the ineffectiveness of this tactic; but, with the ever-rising prominence of data-driven technologies, this case offers a great opportunity for QWE to add another data-oriented process to their repertoire. This opportunity to get in front of customer churn is best described by a quote from the original case document:

> Wall thought that if QWE could estimate the probability that a given customer would leave in the near future and identify the drivers that contributed most to that customers decision, then his team would be able to reach out to the customer, enhance his or her experience with QWE services, and divert churn without giving up costly discounts.

The ability to estimate the probability of churn would enable QWE to positively interact with customers who have not yet decided to unsubscribe but might decide to do so in the future. If QWE continues to pursue a reactive style as their approach to churn, customers will certainly exploit the process in order to get a better deal on their contract even if theyre not seriously considering termination.

Aggrawal (with Wall's advice) came up with a list of potential features (customer age, customer happiness index, and service and usage patterns) that served as first steps in converting the churn handling process into a data-driven process. These features will be described in more detail later in the paper.

1.1 QWE's Goal

The above quotation also sums up Wall's goal for the project. Specifically, QWE wishes to use available data to approximate the probability of customer churn so they can decide which customers should be contacted or offered which incentives to stay on their contract.

First, churn is formally defined; then, proposed are three feature sets to use in a classification model; next, some descriptions are presented to act as the classifier; later, considered is the model within a grand deployment by QWE; finally, expected outcomes and solutions are examined.

2 Churn

This project requires a formal definition of churn.

Definition 1. *For any given time period, churn C is defined as follows:*

$$C = \frac{N_0 - N_1}{N_0}$$

where N_0, N_1 are the numbers of users at the beginning and end of the particular time period, respectively.

It's worth noting that there are many different definitions of churn; however, for this study, a simple definition will suffice. In practice, it would be wise to consider many different definitions both analytically and qualitatively, but it is impossible to numerically examine without actual data.

3 Feature Sets

This paper splits the features into three feature sets:

Baseline Features Features proposed by Aggrawal in the original case

Customer Analytics Features Features pertaining to how customers use the QWE service

Customer Profile Features Features specific to each customer, not necessarily related to the QWE service

Features were split up into sets because they are related to a few different aspects of the customer-to-QWE relationship. Thus, if a feature in one set is more or less effective in a model, then the other features in that set may be more likely to follow suit. This is not verified and is entirely speculative, but it was used because of the fortunate byproduct that it vastly simplifies the process of selecting a model.

One common method for selection is to test models with all possible combinations of features. This is very slow ($m\binom{n}{k}$ training iterations, where m is the number of model hierarchies to test, n is the number of features, and k is the number of features per model), so it would be a computational hindrance to train on all possible feature combinations. Therefore, if N is the number of feature sets to test, we reduce the number of training iterations to $m\binom{N}{k}$, since $N \leq n$. The way these feature sets are defined ensures that $N < n$.

Feature Set	Feature Name
Baseline	Customer Age
	Customer Happiness Index (CHI)
	Number of Support Cases
	Average Support Priority
	Number of Logins
	Number of Blogs
	Number of Views
	Days Since Last Login
Customer Analytics	Contract Type
	Number of First Month Logins
	Rate of Change in Duration Between Logins
	Time Spent on Product Features
Customer Profile	Customer Reviews Sentiment Score
	Frequency of PR Mentions
	Frequency of New Job Postings

Figure 1: Classification of features into feature sets.

Figure 1 shows which features exist in each of the feature sets. Below are corresponding descriptions for each feature grouped by the set.

3.1   Baseline Feature Set

Following are descriptions of each of the features in the baseline feature set. As said above, this set originated from Aggrawal and Wall in the case.

Customer Age Number of months the customer has been subscribed

Customer Happiness Index (CHI) A conglomerate score related to customer experience with a ser-

vice

Number of Support Cases Number of opened support cases per time period

Average Support Priority Average priority of support cases

Number of Logins Number of customer logins per time period

Number of Blogs Number of blog articles written by customer per time period

Number of Views Number of views on blog articles per time period

Days Since Last Login Number of days since customer last logged in

These features all relate to how customers use the QWE product directly. Clearly these are indicators of churn since they influence the customer's opinion of the QWE brand.

3.2   Customer Analytics Feature Set

Following are descriptions of each of the features in the customer analytics feature set.

Contract Type Whether contract is month-by-month or 6-/12-month

Number of First Month Logins Number of customer logins in first month of subscription

Rate of Change in Duration Between Logins The rate of change of the time between login sessions

Time Spent on Product Features Scores on the customer usage of particular product features

The customer analytics features are certainly predictors of churn. The contract type has a temporal influence on churn in the following way: customers who sign up for 6-/12- month contracts are initially committing to a longer relationship with QWE than those who subscribe to month-by-month contracts. This commitment is a somewhat clear reflection of the customer's prior confidence with QWE's service; and, any Bayesian aficionado would argue that prior circumstances influence posterior ones. The remaining three features in this set are descriptors of how customers use the QWE product, which are certainly indicators of churn.

3.3   Customer Profile Feature Set

Following are descriptions of each of the features in the customer profile feature set.

Customer Reviews Sentiment Score Average NLP sentiment score of customer reviews

Frequency of PR Mentions Number of news exposures per unit time

Frequency of New Job Postings The number of new job postings per unit time

This feature set is meant to reflect the nature of the customers themselves as opposed to their activity on the QWE service. It isn't inconceivable that the character of a customer has an influence on their churn. Reviews certainly say something of the public image of the company, which is important because in most cases public image is directly influenced by the company itself. Similarly, the frequency of news mentions affects churn in the same way. Finally, the frequency of new job postings is an indicator of company growth since companies will often have more diverse job postings as they decide to expand.

3.4   Final Notes on Features

The baseline features were accepted because of their suggestion by an industry expert. The customer analytics features expand upon the baseline features while keeping under consideration that the way customers interact with the QWE services affect their decision to churn. Lastly, the customer profile features step away from QWE and look for potential churn patterns in customers based on their corporate profile: it is reasonable to suppose that the qualitative success of customers (companies) is a reliable predictor of churn.

Even with this small handful of features, this project uses the selection method briefly described above: an exhaustive search tactic may be employed to determine which feature sets articulate the best results. However, treating entire sets of features as individuals in the selection process is naive, to say the least. To combat this naivety, principal component analysis (PCA) may be used to extract the most varied features in feature sets while maintaining the ability to filter a fair amount of noise. Upon doing this, meaning in particular features is lost, but general inferences may be extracted from themes of the broader feature sets. Other solutions to the multicollinearity problem include filter, wrapper, and embedded methods; but, using PCA is a good start since more adjustments can be made to the algorithm as results reveal progress.

In any case, these techniques should be experimented with to determine which method derives the best outcomes. Next, the actual model should be briefly considered.

4   Model

If it has not been made clear by now, the model is a classifier. It will output, for any customer, the probability of churn in the next time period. This model could employ anything from logistic regression to a deep neural network. Without actual data, the model is only the product of speculation, which is why this section is so brief. The only requirement is that it predicts the probability of churn for each customer.

5 Deployment

While features and models are important, they are not the overall goal. The broad object of this study is to influence customer churn, not simply to measure it. Thus, this section describes a deployment pipeline that QWE can directly implement into their business operations. Assume at this point that the model has been generated and trained. Let $P(c_i)$ refer to the probability that customer i churns (i.e. output of the model).

1. Collect batch of data

   Since the model is already trained, training data is not used here. Instead, we merely need to collect a new batch of input features for each time period. At the end of the time period, it would be wise to record whether or not the customer actually churned in order to implement a continuous model evaluation.

2. Compute $P(c_i)$

   Use the trained model to compute each customer's probability of churn and pass the results to be segmented.

3. Segment customers into churn categories

   Split the customers based on severity of churn probability. If a customer is more likely to churn, then they should get more attention than a customer who is less likely to churn.

4. Assign support teams to categories and apply incentives

   Use different support teams to treat the different categories of churn. Here, the applied incentives and outcomes (whether the customer churned or not) should be recorded.

5. Examine results

   At the end of each time period, QWE should monitor the metrics that will be described below. Then, they should make corresponding changes to the model/algorithm/deployment as necessary.

QWE can implement this model effectively by following the above steps. It would even be beneficial to periodically retrain the model in order to account for different temporal trends. This would allow the model to "keep up" with current trends. Alternatively, convolutional neural networks can be used to classify sequences of feature values for each customer. However, this is rather advanced compared to the necessary building blocks of the study. In other words, QWE needs to walk before they can run.

6 Evaluation

The only thing left to consider is how QWE can measure their success. This project proposes five evaluation metrics:

1. Overall churn rate

2. Positive/negative service usage

3. Product analytics

4. Effective incentives analytics

5. Long term customer retention

6.1 Overall Churn Rate

The most obvious metric to monitor is the percentage of customers who churn over each time period. This is the same metric that was presented as Definition 1. While the model is predicting churn for each customer, QWE cares about churn overall. Thus, they should evaluate using churn rate as opposed to individual probability of churn.

6.2 Positive/Negative Service and Support Usage

This metric uses the baseline and customer analytics feature sets to create a separate classification model (from hand-annotated data) which monitors the proportion of positive to negative service and support users. Positive users can be characterized as customers who move quickly through different topics in the documentation, engage fully with the products, and interact with their customers whereas negative users might be stuck on certain topics for long periods of time, have trouble with the products and services in general, or outright ignore their customers. Monitoring this ratio before and after different incentive treatments can give us an idea of how the incentives are affecting customers' ability to shift from being negative to positive users, which is a metric QWE should care about.

6.3 Product Analytics

This is not so much a single metric, but instead a continuous project undertaken by an analytics team to gain deeper insights as to how customers use the QWE products and services directly. This is slightly different than service and support usage, which is just a model that results in a ratio to monitor. The analytics

team in charge of this project would be interested in which QWE products and services are being used the most/least and they might investigate the reasons for this as well as how it can be changed.

6.4   Effective Incentives Analytics

This is a metric which can be developed by examining the effects of applying different treatments to even further segmented customer groups. This metric aims to give an idea of which incentives are the most effective for different customer groups. This metric will certainly use the positive/negative service and support user ratio.

6.5   Long Term Customer Retention

Lastly, user retention metrics look at the proportions of users who stay for varying periods of time grouped by different subscription dates. This might allow a better look into the seasonal trends in customer churn: do more customers unsubscribe at some times of the year than others? With this, QWE can see these periodic, long-term trends alongside these time period-dependent metrics. This metric can be visualized easily with a retention matrix.

7   Recommendation & Concluding Statments

This deployment is effective as a whole with continuous attention to the evaluation metrics. First, QWE must pick a time period and gather training data for the main model as well as the classifier that will determine positive/negative service and support usage. Then, they must iterate over the steps presented above over the duration of each time period. This is a continuous process which can be maintained by analytics teams at QWE for as long as management would care to sustain it.