Reducing Customer Attrition with Predictive Analytics

Nate Derby

Stakana Analytics
Seattle, WA

Vancouver SAS Users Group 11/4/15
Introduction: The Problem with Member Attrition

Fisher CU of Ames, IA:

- 6/30/15: 70,359 members, $928MM.
- 3/31/15: 69,534 members, $920MM.
- Grew their membership by 1.19%. That’s great, right?
- Wrong! Those 825 new members are a net gain.
- We actually gained 3541 new members but lost 2716 existing members.

Why is this important?
Introduction: The Problem with Member Attrition

Member attrition dramatically affects net membership growth:

- Retained 20% of departing members:
  ⇒ Gained a net 1368 members, 1.97% net membership growth.

- Retained 25% of departing members:
  ⇒ Gained a net 1504 members, 2.16% net membership growth.

- Retained 50% of departing members:
  ⇒ Gained a net 2183 members, 3.14% net membership growth.

Member attrition has a huge effect!
Introduction: The Problem with Member Attrition

Retaining existing members usually easier and less expensive than gaining new members:

- Members already know and trust you, and you already know so much about them.
- Keeping them might be as simple as making a phone call.
- Key is making full use of our member data.

Focus on members with highest value, highest risk of leaving.
Member Segmentation

- **Divest?**
  - Low Value, High Attrition Risk

- **Aggressively Retain**
  - High Value, High Attrition Risk

- **Cultivate**
  - Low Value, Low Attrition Risk

- **Maintain**
  - High Value, Low Attrition Risk
Many Details to Coordinate!

We’re building a *statistical model*:
- Equation that gives the probability that a member will leave.

Data preparation in three steps:
- Duplicating the data.
- Building our variables.
- Partitioning the data.
What Does This Mean?

We’re building a statistical model:

\[
\text{Probability of leaving in 2-3 months} = f(X_1, X_2, X_3, \ldots)
\]

- \(f\) is a function we don’t know yet (which we’ll build).
- Once we know \(f\), we’ll use \(X_1, X_2, X_3, \ldots\) to get our probability.

In other words ...

- To \textit{build} the model, we need data as of 3 months ago, coupled with who left 2-3 months later (which we know).
- To \textit{use} the model, we need data as of now, to tell us who are likely to leave 2-3 months later (which we don’t know).

The time intervals for \(X_1, X_2, X_3, \ldots\) must be the same for both.
How Do We Do This?

Model Predictor Data Period
(n months)

Model Intervention Period
(1 month)

Model Attrition Observation Period
(2 months)

Actual Predictor Data Period
(n months)

Actual Intervention Period
(1 month)

Actual Attrition Observation Period
(2 months)

December 2010

Now
How Do We Do This in SAS?

SAS Code

```sas
%prepareData( modeling )
%prepareData( scoring )
```
How Do We Do This in SAS?

%%%prepareData Macro

%MACRO prepareData( dataSet );

  %LOCAL now1 now2 now ... attritionEndDate;

  PROC SQL NOPRINT;
  SELECT MAX( effectiveDate )
    INTO :now1
    FROM member_accounts;
  SELECT MIN( tranPostData ), MAX( tranPostDate )
    INTO :startDate, :now2
    FROM member_transactions;
  QUIT;

  %LET now = %SYSFUNC( MIN( &now1, &now2 ) );

Nate Derby
Reducing Credit Union Attrition
How Do We Do This in SAS?

%prepareData Macro

%IF &dataSet = modeling %THEN %DO;
  %LET predictorStartDate = &startDate;
  %LET predictorEndDate = %EVAL( &now - 84 );
  %LET attritionStartDate = %EVAL( &now = 56 + 1 );
  %LET attritionEndDate = &now;
%END;

%ELSE %IF &dataSet = scoring %THEN %DO;
  %LET predictorStartDate = %EVAL( &startDate + 84 );
  %LET predictorEndDate = &now;
%END;

...

%MEND prepareData;
What Are We Doing?

For both of our data sets, we’ll build variables that might be predictive of a member closing his/her account.

- We don’t care if they’re *actually* predictive, as our model will figure that out!
- But we need to “nominate” variables for the model to try out.

Some examples:

- Transaction recency.
- External deposit recency.
- Recent large transaction.
- Small/large number of transactions.
- Seasonality.
How Do We Do This in SAS?

%prepareData Macro

PROC SQL NOPRINT;
  CREATE TABLE predictorData1 AS
  SELECT
    id_member,
    MAX( ( &predictorEndDate - tranPostDate )/7 ) AS tranRecency,
    MEAN( ABS( tranAmt ) ) AS meanTranAmt,
    N( tranAmt ) AS nTrans,
    N( tranAmt )/MAX( INTCK( 'month', tranPostDate, &now, 'c' ) ),
  FROM member_transactions
  WHERE
    tranPostDate BETWEEN &predictorStartDate AND &predictorEndDate
    AND UPCASE( tranTypeCode ) IN ( 'CCC', 'CCD', ... 'WTHD' )
  GROUP BY id_member;

Nate Derby
Reducing Credit Union Attrition
How Do We Do This in SAS?

%prepareData Macro

```sas
CREATE TABLE predictorData2 AS
SELECT
  id_member,
  MAX( ( &now - tranPostDate )/7 ) AS depRecency,
FROM member_transactions
WHERE
  tranPostDate BETWEEN &predictorStartDate AND &predictorEndDate
  AND UPCASE( tranTypeCode ) = 'XDEP'
GROUP BY id_member;
QUIT;
```

Percentiles via **PROC UNIVARIATE**, then merge ...
We won’t build *one* statistical model for our forecasts.

- We’ll build several statistical models.
- We’ll choose the one that gives us the best results.
- We’ll give an estimate of how accurate those results are.

Each of those steps needs a different data set!

- A statistical model finds the equation that best fits the data.
- If we use the same data, then of course we have great accuracy!
- But the whole point is to predict data we haven’t seen yet.

If we never actually tested how well our model predicts unknown data, we could have a nasty surprise.
Much better way:

- We’ll build several statistical models.
- We’ll choose the one that gives us the best results.
- We’ll give an estimate of how accurate those results are.

Each of those steps needs a different data set!

- **Training Set** (60%) = data for building the models.
- **Validation Set** (20%) = data for evaluating results of all models.
- **Test Set** (20%) = data for evaluating results of the final model.
Training vs. Validation vs. Test Set

Build

Statistical Model 1
Statistical Model 2
Statistical Model n

Training Data Set

Assess

Statistical Model 1
Statistical Model 2
Statistical Model n

Validation Data Set

Assess

Final Statistical Model

Test Data Set
How Do We Do This in SAS?

**Training: 60%, Validation: 20%, Test: 40%**

```sas
DATA trainingData validationData testData;
SET inputData;
CALL STREAMINIT( 29 );
randUni = RAND( 'uniform' );
IF randUni < .6 THEN OUTPUT trainingData;
   ELSE IF randUni < .8 THEN OUTPUT validationData;
   ELSE OUTPUT testData;
RUN;
```

From 69,534 members:

- 41,875 in training set (60.22%).
- 13,807 in validation set (19.86%).
- 13,852 in test set (19.92%).
Building a Statistical Model (with different sets of variables)

PROC LOGISTIC DATA=trainingData OUTMODEL=trainingModel1;
  CLASS ageTier( REF='18 and Under' ) / PARAM=ref;
  MODEL attrition( event='1' ) = depRecency ageTier lom
       nProducts calls;
  ODS OUTPUT parameterEstimates = parameter_model1;
RUN;
Assessing a Model

Apply the model to our validation set.

Building a Statistical Model (with different sets of variables)

```sas
PROC LOGISTIC INMODEL=trainingModell;
  SCORE DATA=validationData
    OUT=validationForecasts OUTROC=validationROC;
RUN;
```

validationForecasts and validationROC will be used later.
Assessing a Model

If this is our best model and we want to apply it to our test data set:

Building a Statistical Model (with different sets of variables)

PROC LOGISTIC INMODEL=trainingModel1;
   SCORE DATA=testData OUT=testForecasts OUTROC=testROC;
RUN;

When we’re done and want to make our final forecasts:

Building a Statistical Model (with different sets of variables)

PROC LOGISTIC INMODEL=trainingModel1;
   SCORE DATA=inputData OUT=finalForecasts;
RUN;
Attrition Risk = \frac{1}{1 + \exp(0.07 - 0.0043X_1 - 2.02X_2 - 3.30X_3 - 2.91X_4 - 3.21X_5 - 2.76X_6 - 3.37X_7 - 2.36X_8 - 2.67X_9 - 2.75X_{10} - 4.47X_{11} - 4.17X_{12} + 0.094X_{13} + 4.94X_{14} - 2.99X_{15})}.

- $X_1$ = external deposit recency (weeks).
- $X_2$ = 1 if the member is of age 19-24, 0 otherwise.
- $X_{12}$ = 1 if the member is of age 71 or over, 0 otherwise.
- $X_{13}$ = length of membership (months).
- $X_{14}$ = number of products.
- $X_{15}$ = number of customer service calls in the past month.
For every week that a member goes without an external deposit, his/her odds of attrition multiply by $e^{0.00425} = 1.004$.

The odds of a member aged 19-24 leaving is $e^{2.0201} = 7.5$ times the odds of a member 18 or under leaving.

The odds of a member aged 71+ leaving is $e^{4.1651} = 64.4$ times the odds of a member 18 or under leaving.

For every month that a member continues a membership, his/her odds of attrition multiply by $e^{-0.0938} = 0.91$.

For every product that a member signs up for, his/her odds of attrition multiply by $e^{-4.9355} = 0.0072$.

For every customer service call that a member makes, his/her odds of attrition multiply by $e^{2.9855} = 19.8$.  

# Forecasts

Nate Derby

---

## Fisher CU Member Attrition Forecasts as of 15-06-30.xlsx - Excel

<table>
<thead>
<tr>
<th>Member ID</th>
<th>Risk of Attrition in 2-3 months</th>
<th>Aggregate</th>
<th>Age Tier</th>
<th>Length of Membership (Months)</th>
<th>External Deposit Recency (Weeks)</th>
<th>Number of Products</th>
<th>Customer Service Calls (past month)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>99.99%</td>
<td>$24,219.38</td>
<td>36-40</td>
<td>44</td>
<td>178.01</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>99.99%</td>
<td>$2,868.83</td>
<td>36-40</td>
<td>6</td>
<td>14.11</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>99.99%</td>
<td>$21,795.05</td>
<td>61-65</td>
<td>14</td>
<td>28.66</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>99.99%</td>
<td>$9,474.69</td>
<td>41-45</td>
<td>21</td>
<td>87.23</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>99.99%</td>
<td>$202.41</td>
<td>19-24</td>
<td>16</td>
<td>53.89</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>99.99%</td>
<td>$6,473.96</td>
<td>56-60</td>
<td>24</td>
<td>20.07</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>99.99%</td>
<td>$13,850.80</td>
<td>66-70</td>
<td>22</td>
<td>27.34</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>99.99%</td>
<td>$8,423.36</td>
<td>25-30</td>
<td>24</td>
<td>96.49</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>99.99%</td>
<td>$10,938.19</td>
<td>19-24</td>
<td>32</td>
<td>8.39</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>99.99%</td>
<td>$11,398.00</td>
<td>25-30</td>
<td>56</td>
<td>197.92</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>12</td>
<td>99.99%</td>
<td>$14,239.38</td>
<td>41-45</td>
<td>17</td>
<td>59.12</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>99.99%</td>
<td>$105,157.41</td>
<td>25-30</td>
<td>22</td>
<td>41.49</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>14</td>
<td>99.99%</td>
<td>$17,112.47</td>
<td>46-50</td>
<td>32</td>
<td>140.29</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>15</td>
<td>99.99%</td>
<td>$17,542.99</td>
<td>36-40</td>
<td>64</td>
<td>105.08</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>16</td>
<td>99.99%</td>
<td>$7,555.09</td>
<td>46-50</td>
<td>50</td>
<td>20.18</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>17</td>
<td>99.99%</td>
<td>$12,247.90</td>
<td>46-50</td>
<td>34</td>
<td>91.16</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>18</td>
<td>99.99%</td>
<td>$47,612.51</td>
<td>71+</td>
<td>8</td>
<td>32.31</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>19</td>
<td>99.99%</td>
<td>$11,555.09</td>
<td>31-35</td>
<td>31</td>
<td>126.65</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>99.99%</td>
<td>$13,870.93</td>
<td>56-60</td>
<td>24</td>
<td>2.84</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>21</td>
<td>99.99%</td>
<td>$5,404.15</td>
<td>31-35</td>
<td>34</td>
<td>138.28</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>
This is only half of a retention strategy!

- Nothing will happen without an intervention strategy!
Conclusions

- Predictive analytics can be a powerful tool for member retention.
- These techniques just scratch the surface of how we can reduce member attrition with predictive analytics.
- This approach can also be used to predict other aspects of member behavior.
  - When a member will buy a car.
  - When a member will buy a home.
  - Whether a member is committing a financial crime.
- We can also use these techniques to further cultivate our members.
Further Resources


Nate Derby: nderby@stakana.com