Customer Churn Analysis and Prediction

William Chen
Margaret Ji
Catherine Ruan
Overview

Goals

- Use the consumer dataset to:
  - Segment the Fulton Bank customer base
  - Build a model that predicts customer churn

Agenda

1. Data processing
2. Customer segmentation
3. Feature scoring and predictive model implementation
4. Business recommendations
# Data Processing

## Objective
- Prepare data for analysis by removing and modifying data

## Numerical/Binary
- Keep columns containing relevant characteristics of customer segments

## Categorical
- Find appropriate level of detail
- One-hot encode

## Balances
- Set missing balances to -10,000
- Use smooth symmetric log scaling

## Missing Totals
- Fill blank cells with 0’s or -1’s depending on context

125 Columns
Customer Segmentation

Objective
- Figure out if consumers naturally fall into certain groups

Methods
- Dimensionality reduction
- Finding the optimal number of segments
- Clustering
- Segment analysis

Results
- Churn
- Behaviors
Customer Segmentation

Objective
- Find a more concise representation of data

Method
- Dimensionality reduction
  - Autoencoder
  - Principal Component Analysis
- Performance analysis
  - Reconstruction loss

Percent of Data Described

Reconstructed Data vs. Actual Data

[Graph showing the percent of variance described against dimension]

[Graph showing reconstructed data vs. actual data]
Customer Segmentation

Objective

- Use unsupervised learning to segment customers into groups

Why unsupervised segmentation?

- Cherry picking metrics may not capture nuances in the data
- Unsupervised clustering can cover as much information as possible
- To be understand churn, it is good to first understand its correlation with consumer behavior
- Spectral clustering is best suited for nonconvex geometry

Select best number of clusters based on eigengaps

Perform large scale clustering using KMeans

Examine clustering performance using elbow method
Customer Segmentation

Objective

- Use unsupervised learning to segment customers into groups

Largest increases in eigenvalues

Minimized inner cluster distance
# Customer Segmentation

<table>
<thead>
<tr>
<th>Segment</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Swing by”</td>
<td><strong>Highest churn</strong></td>
</tr>
<tr>
<td>“Loyal”</td>
<td><strong>Lowest churn</strong></td>
</tr>
<tr>
<td>“Techy”</td>
<td>Highest average mobile logins</td>
</tr>
<tr>
<td>“Valuable”</td>
<td>Largest percentage of H-P H-F</td>
</tr>
<tr>
<td>“Active”</td>
<td>Highest average of billpay transactions</td>
</tr>
</tbody>
</table>
Customer Segmentation

**Consumer Segments by Percentage**

**Customer Life Time Value by Segments**

**Churn Rate by Segments**

- Swing by
- Loyal
- Techy
- Valuable
- Active
Customer Segmentation

Income by Segments

Percentage Having Savings Account
Customer Segmentation

- Uber/Lyft Payment Average
- Percentage of Having Direct Deposit Service
## Business Recommendation: Target Customer Segments

### Objective
- Classify customers into segments to provide customers with targeted recommendations that meet their needs and increase loyalty

### Acquisition
- Identify and execute campaigns targeting customers with characteristics of low-churn segments

### Servicing
- Provide services tailored to the customer’s needs based on segment traits
- Recommend or cross-sell products associated with loyalty

### Relationship
- Build in-depth relationships with customers via analytics-based personalized services

### Retention
- Prioritize converting “swing by” customers to other segments with lower churn
# Feature Scoring Procedure

**Objective**
- Which features are giving the most improvements to accuracy in a nonlinear model?

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>125 Choose 2 = 7750 Columns</td>
</tr>
<tr>
<td>2</td>
<td>Fit 7750 Random Forests</td>
</tr>
<tr>
<td></td>
<td>(Each column will be in 124 of the models)</td>
</tr>
<tr>
<td>3</td>
<td>Distribution of average accuracies for each model that a column participated in</td>
</tr>
<tr>
<td>4</td>
<td>Scored column as the equally weighted average of the mean, median, 90th percentile, and max of the distribution</td>
</tr>
<tr>
<td>5</td>
<td>Graphed column scores and picked a natural cutoff point</td>
</tr>
</tbody>
</table>
Feature/Column Scoring Results

Top Scoring Columns

<table>
<thead>
<tr>
<th>Index</th>
<th>name</th>
<th>score</th>
<th>mean</th>
<th>max</th>
<th>90thile</th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>STARTPROD</td>
<td>0.801464</td>
<td>0.876033</td>
<td>0.909008</td>
<td>0.806722</td>
<td>0.880242</td>
</tr>
<tr>
<td>63</td>
<td>TOTAL_ASSETS_FPC</td>
<td>0.747434</td>
<td>0.693626</td>
<td>0.909008</td>
<td>0.66857</td>
<td>0.687631</td>
</tr>
<tr>
<td>82</td>
<td>hascheckingactivity</td>
<td>0.7472</td>
<td>0.701238</td>
<td>0.870022</td>
<td>0.714365</td>
<td>0.695153</td>
</tr>
<tr>
<td>7</td>
<td>NEWPROD</td>
<td>0.747086</td>
<td>0.702956</td>
<td>0.856132</td>
<td>0.73875</td>
<td>0.690867</td>
</tr>
<tr>
<td>43</td>
<td>BROKERAGEBAL</td>
<td>0.743331</td>
<td>0.692753</td>
<td>0.805555</td>
<td>0.698787</td>
<td>0.686809</td>
</tr>
</tbody>
</table>

Bottom Scoring Columns

<table>
<thead>
<tr>
<th>Index</th>
<th>name</th>
<th>score</th>
<th>mean</th>
<th>max</th>
<th>90thile</th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td>102</td>
<td>tot_calls</td>
<td>0.652</td>
<td>0.571116</td>
<td>0.821033</td>
<td>0.649147</td>
<td>0.566702</td>
</tr>
<tr>
<td>10</td>
<td>MOVEDIH</td>
<td>0.653974</td>
<td>0.566805</td>
<td>0.836254</td>
<td>0.648787</td>
<td>0.563971</td>
</tr>
<tr>
<td>53</td>
<td>DEPO_SRV_TOT</td>
<td>0.656639</td>
<td>0.620882</td>
<td>0.74392</td>
<td>0.654359</td>
<td>0.607594</td>
</tr>
<tr>
<td>15</td>
<td>IRACONSFRV</td>
<td>0.657627</td>
<td>0.571757</td>
<td>0.841704</td>
<td>0.649831</td>
<td>0.567124</td>
</tr>
<tr>
<td>16</td>
<td>BROKERAGE SRV</td>
<td>0.657671</td>
<td>0.567384</td>
<td>0.850581</td>
<td>0.648758</td>
<td>0.563961</td>
</tr>
</tbody>
</table>
### Most vs. Least Likely to Churn

<table>
<thead>
<tr>
<th>STARTPROD</th>
<th>% churn</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITAC</td>
<td>1</td>
<td>99</td>
</tr>
<tr>
<td>TUNA</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>IRAC</td>
<td>0.931148</td>
<td>305</td>
</tr>
<tr>
<td>OLB</td>
<td>0.916667</td>
<td>12</td>
</tr>
<tr>
<td>CDPB</td>
<td>0.888092</td>
<td>2091</td>
</tr>
<tr>
<td>MMPer</td>
<td>0.846154</td>
<td>2954</td>
</tr>
<tr>
<td>FTAC</td>
<td>0.843844</td>
<td>333</td>
</tr>
<tr>
<td>LCIND</td>
<td>0.814592</td>
<td>12895</td>
</tr>
<tr>
<td>MTGS</td>
<td>0.79519</td>
<td>7734</td>
</tr>
<tr>
<td>EOQPT</td>
<td>0.795181</td>
<td>581</td>
</tr>
<tr>
<td>EOQMTG</td>
<td>0.782609</td>
<td>92</td>
</tr>
<tr>
<td>BUNP</td>
<td>0.75</td>
<td>12</td>
</tr>
<tr>
<td>CKINT</td>
<td>0.74317</td>
<td>22073</td>
</tr>
<tr>
<td>TB</td>
<td>0.736864</td>
<td>2398</td>
</tr>
</tbody>
</table>

### Churn Fraction by Product

The churn fraction by sorted product index is shown in the graph. The churn fraction is calculated as the proportion of users who churn within a given period of time. The graph displays the churn fraction for each product, sorted by the index of the churn fraction. The x-axis represents the index of sorted churn fraction, while the y-axis represents the churn fraction. The data suggests that some products have a higher churn rate than others, with products like FILN, NLON, and A having significantly higher churn fractions compared to products like VD, NILN, and ND.
Predictive Model Implementation

Objective

- Design a model that predicts whether a household churns or is kept

Method Implementation Notes

- Nonlinear model allows for complex interaction
- By using “class_weight = ‘balanced’” in the model, we make sure the accuracy on the kept households and churned households are prioritized equally
- Since there are fewer churned HH, the precision suffers, but this is in line with business intuition of losing a customer is more expensive than the cost to keep an existing customer from churning
Random forest “sqrt” vs “all” refers to checking sqrt(features) or all features at each split
- F.S means the model is run on feature selected data - the top 35 rows
- Logistic regression has poor performance, but we may be able to get more meaningful significance data out of it.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy-Kept HH</th>
<th>Accuracy-Churned HH</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest (sqrt)</td>
<td>0.867</td>
<td>0.94</td>
<td>0.656</td>
<td>0.94</td>
<td>0.773</td>
</tr>
<tr>
<td>Logistic Regression w/ L2</td>
<td>0.408</td>
<td>0.848</td>
<td>0.28</td>
<td>0.848</td>
<td>0.421</td>
</tr>
<tr>
<td>Random Forest (all)</td>
<td>0.792</td>
<td>0.991</td>
<td>0.564</td>
<td>0.991</td>
<td>0.719</td>
</tr>
<tr>
<td>F.S. Random Forest (sqrt)</td>
<td>0.866</td>
<td>0.933</td>
<td>0.655</td>
<td>0.933</td>
<td>0.769</td>
</tr>
<tr>
<td>F.S. Random Forest (all)</td>
<td>0.793</td>
<td>0.991</td>
<td>0.565</td>
<td>0.991</td>
<td>0.72</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>0.988</td>
<td>0.92</td>
<td>0.954</td>
<td>0.92</td>
<td>0.936</td>
</tr>
<tr>
<td>F.S. AdaBoost</td>
<td>0.983</td>
<td>0.922</td>
<td>0.935</td>
<td>0.922</td>
<td>0.928</td>
</tr>
</tbody>
</table>
## Business Recommendation: Utilize Predictive Variables

<table>
<thead>
<tr>
<th>Objective</th>
<th>Understand</th>
<th>Predict</th>
<th>Strategize</th>
</tr>
</thead>
</table>
| • Use variables most predictive of churn to inform insights and strategies personalized for the customer | • Examine intuition for starting product and other high-scoring variables  
• Improve data tracking to include more of customers product profile | • Ensure data is robust enough to draw conclusions  
• Use a nonlinear model to predict whether customers are likely to churn | • Reconsider profitability of high-churn products  
• Encourage customers to switch to or add low-churn products |
## Conclusion

### Goals
- Use the consumer dataset
- Segment the Fulton Bank customer base
- Build a model that predicts customer churn

### Recommendations

- **Customer segmentation**
  - Use customer characteristics to segment customers, allowing for easier acquisition, servicing, relationship development, and retention of customers

- **Predictive model**
  - Predict the likelihood of churn in an individual customer
  - Formulate strategies based on a trait’s association with high or low churn