Propensity Modeling of Handset Upgrades

A real time class project with ABC Wireless
Faculty Advisor: Dr. Goutam Chakraborty
Introduction

- **Project objective:**
  Profile and differentiate customers upgrading to a handset with ABC from those who are not in the existing database
  Predict whether or not, they would upgrade to a smart phone
- **Phase 1:**
  Data Audit and Descriptive Analysis of 3 million observations and 90 variables in the Customer database
  Observe the data patterns and identify key relationships between input and target variables
- **Phase 2:**
  Build predictive models and analyze interaction effects. Choose optimized model for scoring
Data Audit and Predictive Analysis

PHASE-I

Variable inspection using Cross-tabs, Summary statistics and distributions

- Delete observations with max variables missing
- Handle missing values
- Handle outliers
- Recode the categorical variables or reduce the levels
- Create new variables as needed
- Draw a balanced sample
- Force the significant variables in managerial sense
- Drop the insignificant variables in the statistical sense
- Evaluate the relative variable importance

PHASE-II

- Use the Input variables from Phase-I
- Use variable selection (if needed)
- Transform the variables (if skewed)
- Build various models using various options in SAS EM
- Compare and Evaluate Models
- Choose an Optimized model to achieve max prediction on prospects

Various Tests and Correlation Analysis
Data Preparation

- Eight categorical variables were recoded and 23 variables’ missing values were imputed with relevant values.
- Four new variables were created to aid analysis.
- Outliers and extreme values were eliminated from five interval variables and suitable transformations were considered for eight others.
- 17% of the data had missing values for most of the variables. These records were subset for separate analysis.
- A 30000 sample was drawn for consideration.
- 35 significant variables were identified.
- The relative importance of the variables was analyzed.
Variable worth & Interaction effects

- Variable worth and Chi Square

Interaction effects of variables two at a time

<table>
<thead>
<tr>
<th>First Variable</th>
<th>Second Variable</th>
<th>Max-Rescaled R Squared Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue</td>
<td>_2_Time</td>
<td>0.0203</td>
</tr>
<tr>
<td>Current_Rev</td>
<td>Active_Phone_Count</td>
<td>0.025</td>
</tr>
<tr>
<td>Current_Rev</td>
<td>Mou_Current_Normalized</td>
<td>0.0476</td>
</tr>
<tr>
<td>Current_Rev</td>
<td>Num_features</td>
<td>0.0244</td>
</tr>
<tr>
<td>Active_Phone_Count</td>
<td>Credit_90</td>
<td>0.0047</td>
</tr>
<tr>
<td>Active_Phone_Count</td>
<td>Mou_Current_Normalized</td>
<td>0.0359</td>
</tr>
<tr>
<td>Active_Phone_Count</td>
<td>Num_features</td>
<td>0.0207</td>
</tr>
<tr>
<td>Active_Phone_Count</td>
<td>_1_Time</td>
<td>0.008</td>
</tr>
<tr>
<td>Credit_90</td>
<td>_2_Time</td>
<td>0.0014</td>
</tr>
<tr>
<td>Mou_Current_Normalized</td>
<td>_1_Time</td>
<td>0.0371</td>
</tr>
<tr>
<td>Mou_Current_Normalized</td>
<td>_6_Time</td>
<td>0.0323</td>
</tr>
<tr>
<td>_1_Time</td>
<td>_2_Time</td>
<td>0.0049</td>
</tr>
<tr>
<td>_1_Time</td>
<td>_6_Time</td>
<td>0.0107</td>
</tr>
<tr>
<td>_2_Time</td>
<td>_6_Time</td>
<td>0.0015</td>
</tr>
</tbody>
</table>
Preliminary analysis results

Most of them upgrade with handset types B or C

Most of the upgrades take place in first two quarters of 2010

98.06% of the customers in the sample who upgraded to Type ’B’ are Voice Customers and not payGo.
Predictive Models

- With 32 input, one ID and two target variables, various models were built and evaluated for optimized fit statistics
- Further replacement, imputation and transformation options were deployed
- 50% of the data was used for training and the rest was used for validation
- Neural Networks, Regression, Decision Trees and Ensemble models were compared against each other. Different options in SAS Enterprise Miner were explored
- Binary Target was modeled for least validation misclassification rate and the Nominal Target was modeled for the maximum ROC Index as the statistic
Fit Statistics

- Ensemble model built with the average predictions of neural network, regression and decision tree outperforms other models in with the maximum validation ROC Index of 0.743.

- The lift of the various models built is illustrated above.
Results and Recommendations

- The regression model performing a stepwise selection with a probability of 0.05 to enter and stay in the model and with validation misclassification rate as the statistic outperforms other Neural Network Architectures like RBF, MLP and decision tree purity parameters like Gini and Entropy.

- The effort to build more optimized model is still on.
- Recommendations to ABC Wireless were documented.
Conclusion

- Data cleaning and manipulation of the customer database using various SAS procedures and inbuilt options was performed.
- Various predictive models were built and analyzed for relative significance with respect to the selected fit statistics.