Direct Marketing Optimization

Quissuiven (Qui) Tai

Agenda

- Data Processing
- Feature Engineering
- Preparing training and test sets
- Model training and evaluation
- Revenue calculation

Data Processing

- Dealing with missing values:
  - Sex variable: replace with “unknown”, since not fair to impute with most frequent value
  - Other variables: replace with 0, which makes sense since they refer to number of financial products, transactions and account balance
- Label encoding of categorical variables:
  - There is only 1 categorical variable – Sex. The categorical values are converted to numeric values using label encoding.

Feature Engineering

6 new features were introduced for propensity models of credit products

- **Total Credit Product Count**: total number of credit products such as live credit cards, overdrafts and consumer loans
- **Credit Product Balance**: total amount of liability/debt for credit products
- **Credit vs Debit Turnover**: ratio of credit turnover to debit turnover
- **Credit Debit Transaction Ratio**: ratio of number of credit transactions to debit transactions
- **Credit Transaction Frequency**: average number of credit transactions per month
- **Debit Transaction Frequency:** average number of debit transactions per month

Preparing training and test sets

- The 969 samples with existing labels are used to define the training and test sets. The remaining 646 samples are set aside for inference using the trained model later.
- Stratified 70:30 split to ensure sufficient test samples to evaluate performance
- Upsampling of minority class using SMOTE to deal with class imbalance. This showed better performance compared to giving higher weight to minority class during training.

Model Training and Evaluation

- Several tree-based models such as Random Forest, LightGBM and XGBoost were fitted on the training set and evaluated on the test set
    - Tree-based models were selected since they capture non-linearities in tabular data well and they are robust to multicollinearity.
    - Evaluation metric: F1 score
- Model performance was optimized by hyperparameter tuning using randomized search.

Model Training and Evaluation: Results

Consumer Loan

	precision	recall	f1-score	support
0.0	0.75	0.80	0.78	204
1.0	0.45	0.38	0.41	87
accuracy			0.68	291
macro avg	0.60	0.59	0.59	291
weighted avg	0.66	0.68	0.67	291

Top 5 Features:
1. ActBal_CA
2. Age
3. Tenure
4. Credit_vs_Debit_Turnover
5. Credit Transaction Frequency

Credit Card

	precision	recall	f1-score	support
0.0	0.78	0.88	0.82	218
1.0	0.40	0.25	0.31	73
accuracy			0.72	291
macro avg	0.59	0.56	0.56	291
weighted avg	0.68	0.72	0.69	291

Top 5 Features:
1. ActBal_CA
2. Tenure
3. Age
4. Credit_vs_Debit_Turnover
5. VolumeCred

Mutual Fund

	precision	recall	f1-score	support
0.0	0.82	0.89	0.85	233
1.0	0.32	0.21	0.25	58
accuracy			0.76	291
macro avg	0.57	0.55	0.55	291
weighted avg	0.72	0.76	0.73	291

Top 5 Features:
1. ActBal_CA
2. Tenure
3. Age
4. VolumeCred
5. VolumeCred_CA

Revenue Calculation

Given the 3 propensity models,

1. Calculate median revenue for each product
2. For each client, take propensity to purchase * median revenue to get expected revenue for each of the 3 products
3. Take the maximum expected revenue across the 3 products to decide which offer the client should be targeted with
4. Select the top 15% of clients in terms of expected revenue to be targeted

Final expected revenue is 2323.10, where all the 242 clients are targeted with Consumer Loan offer.