Executive Summmary: Direct Marketing Optimization

Bugnot Reinelle Jan Cruz
Data Scientist
reinbugnot@gmail.com

I. Introduction

This case study leverages machine learning models to identify high-propensity clients for three financial products, Consumer Loans (CL), Credit Cards (CC), and Mutual Funds (MF), and develop a targeted marketing strategy for each. Specifically, the objectives are as follows:

1. Predict client purchase propensity for different financial products.
2. Optimize the marketing campaign by selecting the best clients to target.
3. Maximize revenue potential while adhering to specified outreach limitations.

II. Methodology

A. Exploratory Data Analysis

The raw dataset consisted of four sub-datasets:

- *Soc_Dem*: socio-demographic data
- *Products_ActBalance*: account balances
- *Inflow_Outflow*: financial transaction records
- *Sales_Revenue*: sales and revenue data

The first three sheets contained input features, while the last represented the target variable. To ensure data reliability and consistency, the following preprocessing steps were taken:

1. **Duplicate Check**: Verified that no duplicate records existed across any dataset.
2. **Missing Value Treatment**: Identified 28 clients classified as dormant (present in demographic and product datasets but missing from transaction records). These entries were retained, and NaN values were zero-filled to reflect their dormant status.
3. **Data Type Validation**: Label-encoded the Sex attribute, the only categorical variable.
4. **Outlier Detection**: Identified heavy right-skew in financial attributes and applied Winsorization at the 95th percentile to cap extreme values.
5. **Feature Engineering**: Created three new features to enhance model performance:
    a. "Dormant": Flag indicating inactive accounts.
    b. "Total_ActBal": Aggregated balance across all accounts.
    c. "Credit_Debit_Ratio": Ratio of credit inflow to debit outflow.
6. **Redundant Feature Removal**: Detected high correlation between adjacent transaction features and removed redundant current account (*_CA*) variables to reduce multicollinearity.
7. **Target Class Imbalance**: Observed significant class imbalance in target variables, which was later addressed through class weighting and resampling.

B. Model Configurations

To predict purchase likelihood across the three financial products, I implemented and evaluated three LightGBM models per product:

- Manually tuned LGBM with class weighting
- Manually tuned LGBM with SMOTE resampling for class balance
- LGBM with 5-fold cross-validation and class weighting, optimized using *Optuna*

Optuna is an automated hyperparameter optimization framework based on the Bayesian optimization method, that allows the automatic searching and pruning of hyperparameters across a given parameter space. Using the Optuna framework, I can quickly find optimal parameters that maximizes the performance of the LGBM model.

C. Evaluation Criteria

The evaluation metric I used to optimize the models is a composite metric between PR-AUC Score and Brier Loss:

$$score = pr\_auc - (brier\ loss \times 0.1)$$

This metric ensured that the selected model maximized Precision-Recall AUC to effectively rank positive cases—i.e., successful sales—in imbalanced datasets where positive instances are relatively rare, while also maintaining well-calibrated probabilities through the minimization of Brier loss.

D. Benchmark Results

Table I shows the benchmark results across the three model configurations, specified in Sec. II, B, and the three financial products, specified in Sec. I.

TABLE I. Benchmark Results

	base	**base_smote**	**optuna_lgbm**
MF	0.230830	0.220438	**0.249838**
CC	0.389257	**0.395533**	0.355356
CL	0.391780	0.411490	**0.439773**

The top performing models (as highlighted) for each product category will be selected and used in the subsequent steps in the case study.

III. Revenue Optimization

A. Expected Revenue

To optimize revenue generation under the provided marketing constraints, I leveraged the expected value formula as a guiding principle:

$$E[X] = P(Y) \times E[X|Y] \quad (1)$$

where:
$E[X]$: the expected revenue
$P(Y)$: the sale predicted probability (i.e. propensity)
$E[X|Y]$: the average revenue when a sale occurs

B. Propensity

The selected models for each product were used to generate prediction probabilities across the entire dataset. These probability values represent the propensity of the client to purchase the product. Fig.1 shows the resulting distribution of propensity values across the models for all 3 products.

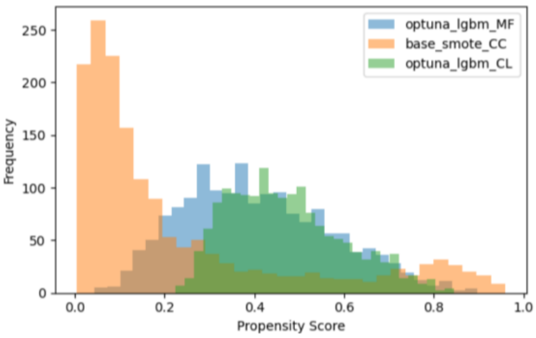

Fig. 1. Distribution of predicted propensity values per model

The distribution shows the differences in how the models assess client purchase likelihoods. The CC model (orange) is right-skewed, with its tail extending toward higher probability values, indicating a tendency to assign lower scores to most clients while reserving very high scores for a few. In contrast, the MF (blue) and CL (green) models follow a more normal distribution.

C. Average Revenue per Product

Finally, the average revenue per product is determined by computing the mean revenue from all successful sales of each product, as shown in Table II.

TABLE II. Average Revenue per Product ($, Sale Price)

MF	CC	CL
$9.66	$10.86	$12.04

Using Eq. (1), I evaluated two targeting strategies:

1. Pure Revenue Maximization: Selecting the top 15% of clients solely based on expected revenue.
2. Balanced Optimization: Prioritizing high-propensity matches first, then ranking by revenue.

Both methods were effective, but pure revenue maximization achieved a slightly higher expected return (**S$1,359** vs. **S$1,330**). While the balanced approach ensured more diverse product distribution, the pure revenue method aligned more directly with the primary objective of revenue maximization.

IV. Insights and Discussion

By examining the mean and median values of key demographic and financial data for clients with the highest propensities, we can identify the defining characteristics of the ideal target clients for each product:

TABLE III. Target Client Characteristics per Product

Consumer Loan (CL) 'Average Age': 23.62, 'Median Age': 24.0, 'Sex Distribution': {1: 0.5, 0: 0.5}, 'Average Tenure (Months)': 183.19, 'Median Tenure (Months)': 179.0, 'Average Account Bal': 5158.33, 'Median Account Bal': 2071.90, 'Average C/D ratio': 1.84, 'Median C/D ratio': 0.95, 'Dormant Distribution': {False: 1.0}}	Early adults around 24 years old who have been with the bank for a very long time (approx. 12 years) which is unusual given their relatively young age. They are likely clients who may have been associated with the bank through joint accounts with parents. These clients are now entering major life stages that require financial support, such as education, early career investments, or first-time asset purchases.
Credit Cards (CC) 'Average Age': 44.34, 'Median Age': 42.0, 'Sex Distribution': {1: 0.53, 0: 0.47}, 'Average Tenure (Months)': 104.07, 'Median Tenure (Months)': 95.0, 'Average Account Bal': 12119.72, 'Median Account Bal': 10632.11, 'Average C/D ratio': 0.96, 'Median C/D ratio': 0.95, 'Dormant Distribution': {False: 0.99, True: 0.01}	Middle-aged individuals. They have been with the bank for about 8.7 years on average and maintain substantial account balances (median of ~$10.6k) and close-to-1 credit-to-debit ratios (average of ~0.96). These suggest they may have strong financial management habits, likely paying off their credit card balances regularly while maintaining liquidity. These clients may be prime candidates for credit card offerings.
Mutual Funds (MF) 'Average Age': 42.98, 'Median Age': 40.5, 'Sex Distribution': {1: 0.57, 0: 0.43}, 'Average Tenure (Months)': 106.64, 'Median Tenure (Months)': 103.0, 'Average Account Bal': 4441.28, 'Median Account Bal': 1118.60, 'Average C/D ratio': 72.78, 'Median C/D ratio': 0.9844310178617752, 'Dormant Distribution': {False: 1.0}	Middle-aged individuals with a slight male majority (57%). They have been with the bank for approximately 8.9 years and maintain moderate account balances (median of ~$1.1k). The high average credit-to-debit ratio of 72.79 suggests majority of these clients heavily rely on credit products or engages in frequent large transactions. Their consistent account activity and relatively stable financial engagement suggest they are financially aware and actively managing their investments.