Client Segmentation and Product Recommendation

A Machine Learning Framework for Bank Decision Support

Luca Amadori, Margherita Bencini, Andrea Catelli, Rodolfo Coppola, Stefano Corti

Introduction

In a competitive and highly regulated financial environment, banks must align their product offerings with the specific needs and risk appetites of their clients. This project addresses the challenge by **developing a comprehensive data-driven pipeline for client segmentation and personalized product recommendation**.
Using a customer dataset that includes financial and demographic information, our objective is to forecast each client's propensity to invest in income-oriented or accumulation-oriented products. The ultimate goal is to improve customer satisfaction, optimize product allocation, and improve business performance through targeted financial advice.

Methodology

We begin with a detailed exploratory data analysis to understand client characteristics and product distribution, applying transformations (Box-Cox, logarithmic) to normalize skewed variables like Income and Wealth. The predictive modeling phase focuses on classifying clients based on their likelihood to adopt Income or Accumulation investments. We employ several machine learning models: the best are Random Forests, Artificial Neural Networks, and XGBoost, evaluated against our **baseline logistic regression**. To boost performance, we incorporate **feature engineering** (e.g., income-to-wealth ratio, financial education-to-age ratio) and **hyperparameter tuning** (via grid search or Optuna).

In parallel, we analyze the bank's current product catalogue and identify gaps in the risk spectrum, particularly in low- and high-risk segments. To address these, we **create new financial products using synthetic replication** via futures contracts, ensuring regulatory compliance by computing risk levels through the SRRI (Synthetic Risk and Reward Indicator) in accordance with CESR guidelines. For risk estimation, we apply the Cornish-Fisher expansion to account for non-normal return distributions.

Main Findings

Our best predictive models achieve strong performance: for accumulation strategies, **XGBoost** reaches an **F1-score** of **0.828**, while for income strategies, it attains **0.661**. **Ensemble models** combining XGBoost, Neural Networks, and Random Forests offer additional robustness but occasionally trade off recall for precision. The product creation phase successfully introduces new low-risk time deposits and a synthetic accumulation product that becomes the most frequently recommended item. However, we were unable to replicate a satisfactory high-risk income product due to data limitations.

Discussion

To bridge model (machine learning) evaluation metrics and financial outcomes, we created the *Business Impact Score* (BIS), a profitability-oriented metric that captures **potential revenue from correct recommendations and penalizes misclassifications**. BIS is based on the idea that for a bank, clients are not all equal: higher wealth clients impact the bank more. To do so, we did some industry research and associated some likely levels of commissions for each type of product, based on the description and the risk level. We were then able to define our metric as:

$$BIS = \frac{1}{N}(\sum_{TP}(comm \cdot wealth) - \alpha \cdot \sum_{FN}(\overline{comm} \cdot wealth) - \beta \cdot \sum_{FP}(\overline{comm} \cdot wealth))$$

BIS is measuring the average per capita profit from product commissions. Correct predictions (true positives) are considered entirely, considering the actual realized commissions. Missed gains (false negatives) are estimated with the average of commission per wealth, and weighted for a factor $\alpha = 1$. For incorrect predictions (false positives) we penalize for annoying a client with a product that they don't desire. This is taken into account by the coefficient $\beta = 0.1$.

BIS analysis confirms **XGBoost** as the most profitable model in both types of investment. This integration of modeling accuracy with business performance offers practical value, helping financial institutions prioritize both technical performance and commercial outcomes.

The recommendation system, built on top of the classification models, assigns products based on **predicted investment behavior** and **individual risk tolerance**. The addition of our new products significantly improves coverage, enabling us to assign suitable offerings to nearly all clients.

Conclusions

This project demonstrates how machine learning can **improve financial decision** making by linking client profiling, predictive analytics, and synthetic product design. The pipeline not only predicts client behavior, but also creates new investment solutions to serve previously under-served segments. The introduction of the Business Impact Score adds a critical dimension, aligning technical success with **business profitability**. Future work may focus on developing higher-risk products, which would ensure higher commissions to the bank as well, perhaps expanding our index dataset.