The Predictive Value of Internal App Data as an Alternative to Open Banking

Francisco António Mendonca
Javier Ocáriz Gallego

Revolut

Context

Data availability and alternatives to open banking data

- A primary challenge in credit modeling arises from **data gaps**, where specific variables are omitted by design during the underwriting process. Here it is investigated a specific manifestation of this data heterogeneity: the absence of open banking data for customers who proceed through a particular underwriting funnel;
- Build credit scoring models for customer funnel's yet to be deployed;
- Alternatives to open banking are explored, including using transaction features built from data captured in Revolut's mobile application;

Our approach: Build a unified ML model using forced splits

- **Tree boosting model using a forced split criteria**
  - **Problem:** Revolut plans to start lending through a new customer funnel, which does not require capturing open banking data. No data from this funnel exists;
  - **Approach:** Duplicate the data for personal loans and set all open banking features to NULL. This ensures the target distribution remains unchanged as well as the correlations between remaining features;
  - **Mechanism:** With the duplicated dataset, construct a boolean flag indicating whether Open Banking data is present or not. Use this flag to force split each tree in the boosting sequence. The optimization process begins after this split.
  - **Benefits:** Allows to score customers in both funnels with a single model, thus reducing the maintenance burden. In each submodel (i.e., after the initial forced split), features are allowed to have different relative importance as well as splitting criterias.

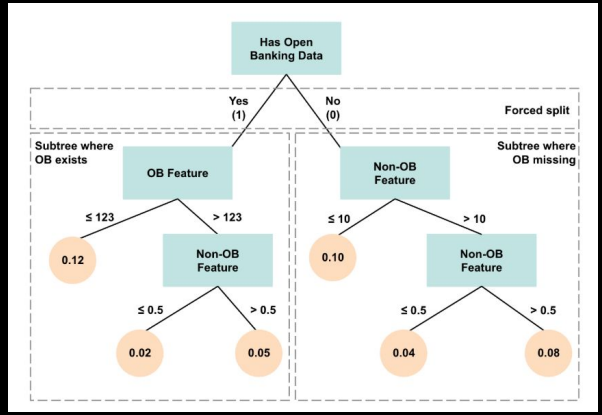
Methodology

Data

- Because no data for the funnel with missing open banking exists, the first step consists in duplicating the entire dataset for customers that applied to personal loans. This ensures the feature distributions are kept unchanged (except for open banking as per next step) as well as the dependent variable;
- Set all open banking features to **missing.** This will mimic the new funnel yet to be deployed.
- Assign a boolean flag indicating whether open banking data was assigned to **missing**

Model design

- The model is built using a LightGBM framework. Construct the forced split .json file by indicating the index of the splitting feature within the list of features as well as the value for splitting;
- Tune model hyperparameters using grid search and feature selection using forward selection with SHAP values;
- Train the final model and evaluate whether the forced split criteria is being enforced;
- Evaluate the model globally and on each customer funnel.



Contribution of open banking data

- After the final model is evaluated, comparing model performance between each customer funnel, allows for a direct assessment of open banking contribution;
- Between funnels, due to the greedy splitting, each feature can have different importances (rank) and be split on different thresholds. The result is a model that tried to minimize the prediction error conditional on the available data.

Revolut

Business case: Application to French and Spanish Market

Data

- **Sample:** For each market, data for personal loans and credit cards is extracted. After preprocessing the data (duplicating and assigning open banking features to NULL), the resulting datasets contain 16 626 and 8 297 observations for France and Spain, respectively.
- **Window:** The data spans from 2022 onwards with performance tracked up to 12 months after origination.
- **Features:** Uses Deep Feature Synthesis to create features from relational tabular data.

Market	Observations
France	16 626
Spain	8 297

Source	Description
Application data	Information provided by the customer at the time of onboarding, including basic demographic and economic declaration
Past credit history	Information about past credit applications and customer credit performance
Internal app usage	Information related to customers spending/income/savings/investment behavior, recorded in Revolut's mobile app
Other internal data	Other type of data excluded from previous sources, existing within Revolut's database. Examples include device information, other declared information and similar attributes
Open banking data	Data obtained after customer links external accounts. Reflects customer spending/income habits
Bureau data	Data obtained from the official credit bureaus in each market

Business case: Application to French and Spanish Market

Results

- **Evaluation metric:** Using the AUC as the performance metric, globally and within each customer funnel, **allows to directly evaluate the contribution of open banking that cannot be extracted from the remaining features**
- **Feature importance and effect:** Using SHAP values, feature relative importance and effect is assessed over each customer funnel.

Key findings

- **Feature importance:** In the absence of open banking data, internal Revolut app data takes the most important role in driving model performance. In instances where open banking features are absent, the model's learning process infers distinct relationships between the remaining available features and the target variable. Consequently, feature importances and splitting criteria within each subtree undergo unrestricted optimization. In addition, because the splitting criteria changes from funnel to funnel, it is possible to see in the SHAP plots how the missing open banking data impacts the magnitude of the effects that each feature has in driving predictions.
- **Impact of open banking data:** The results show that, when available, open banking features rank relatively high and offer a modest increase in predictive power. However, for the path where no open banking exists the model uses the remaining features differently. The submodel without open banking data attains a comparable rank ordering performance to the model leveraging open banking data.

Difference in AUC between funnels		
Market	**France**	**Spain**
Train	+2.78	+4.0
Test	+3.26	+0.5

Thank you

Revolut