Enhancing Credit Risk Models at Revolut by Combining Deep Feature Synthesis and Marginal Information Value

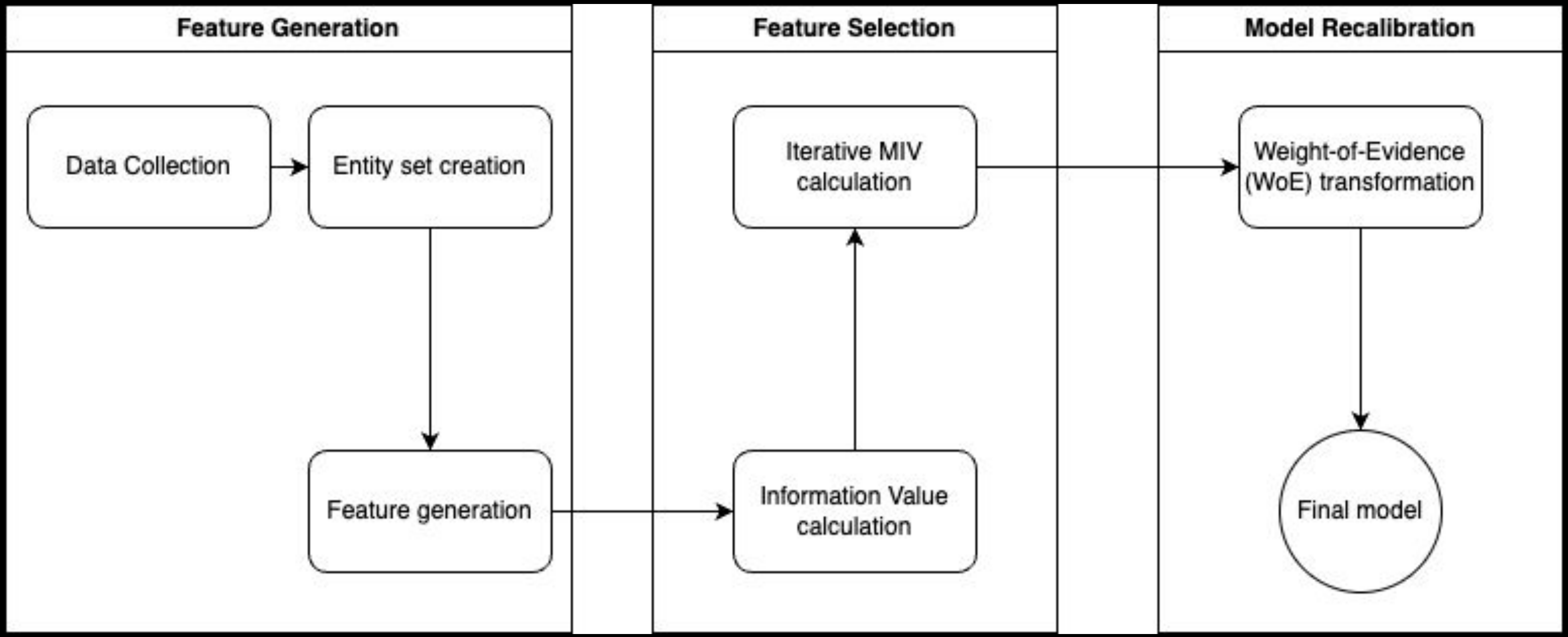
Federico Spinella & Tadas Krisciunas
Edinburgh Credit Risk Conference
2025-08-29

Revolut

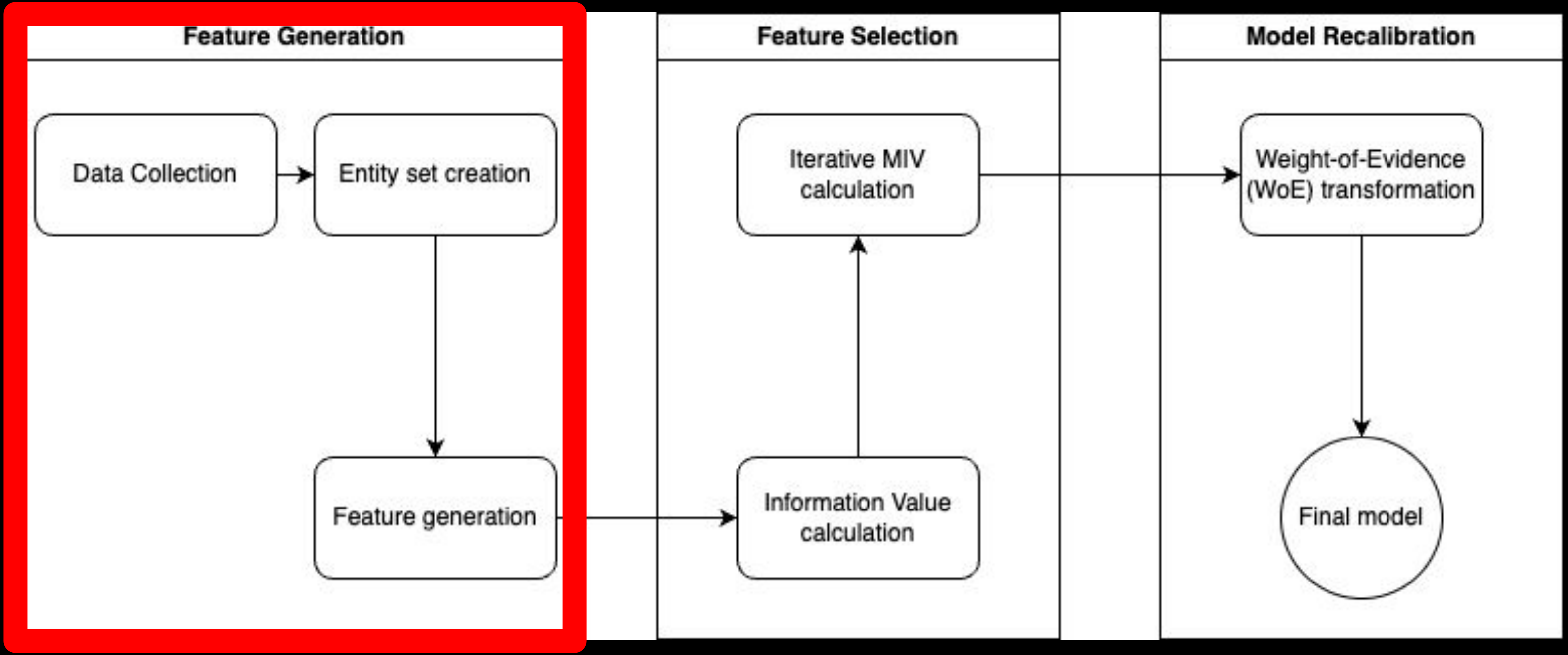
Acquisition Probability of Default Model Development Methodology

Introduction and Overview

PD Model Methodology: Overview



PD Model Methodology: Feature Generation



Acquisition Probability of Default Model Development Methodology

Feature Generation

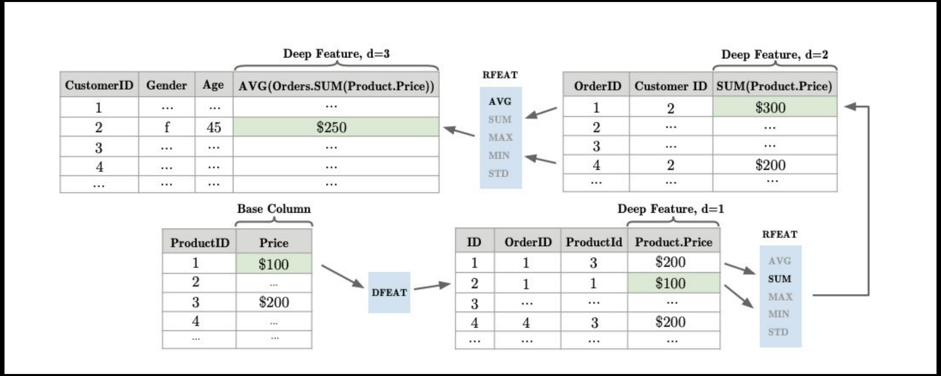
Problem

In the domain of credit risk modelling, the generation of predictive features from complex, relational datasets is paramount for accurate risk assessment. Traditional methods often involve manual feature engineering, which can be time-consuming and may not capture intricate relationships within the data.

Solution

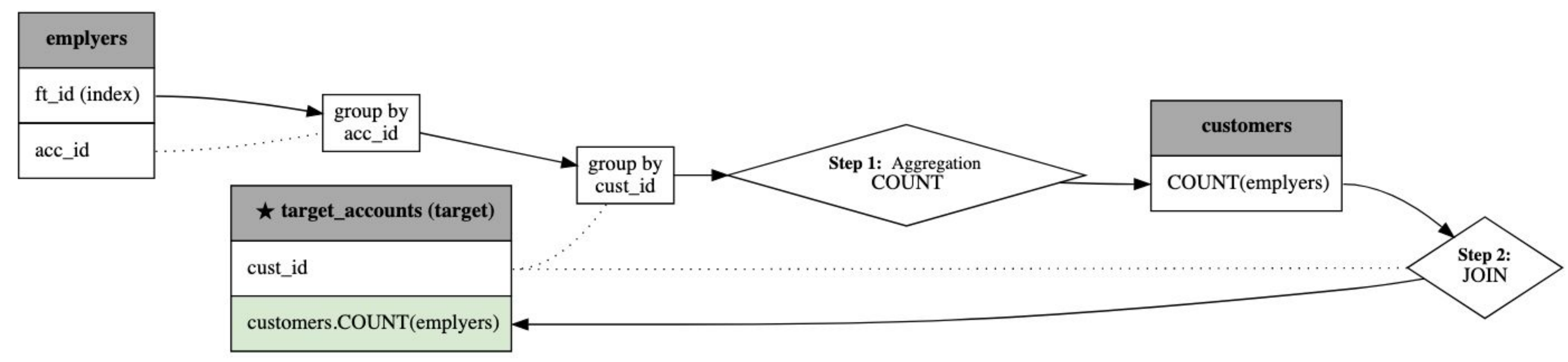
Based on a MIT paper, Deep Feature Synthesis (DFS) is an algorithm that automatically generates new features based on the relational and temporal structure of the data.

The Deep Feature Synthesis (DFS) algorithm is designed to automatically generate features for relational datasets. This algorithm operates by traversing the relationships within the data, moving towards a base field, and then applying mathematical functions sequentially along that path to construct the final feature. The sequential application of calculations allows for the definition of features with a certain "depth," hence the name Deep Feature Synthesis.



Revolut

Example 1.



Example 2.

```
customer.TREND(account_history.amount_past_due, last_updated_on WHERE account_history.contract_type = D3.Personal_Loan)
```

Target Entity | Aggregation function | Attribute | Timestamp | Filter

Revolut

Automation: Transactional Data

All transaction-based features based on Deep Feature Synthesis are **automatically calculated for all credit applications** from the previous day nightly to make them readily available for model development and monitoring.

We picked transactional data first as it is both **highly predictive** of credit risk and **does not vary country-to-country**. Credit bureau/history data varies by country and needs separate pipelines/logic in each country.

1. **Fetch Application Data**
   Fetch application data and split it into fixed sized batches
2. **Fetch Transaction Data**
   Fetch transaction data for each type
3. **Reconstruct Transaction Data**
   Reconstruct extracted data to state of *context_date*
4. **Convert currency to GBP**
   Use *Currency Converter* to convert all tx to GBP
5. **Split Datasets by tx State**
   Split transactions data based on *transactions_state*
6. **Create Balance Dataset**
   Create Balance dataset from tx datasets with *transaction_state == COMPLETED*
7. **Calculate Features**
   Load data into predefined *FTEntity* instances and initialize *ft.EntitySet* which is used to calculate features

Acquisition Probability of Default Model Development Methodology

Weight of Evidence, Information Value and Marginal Information Value

Weight of Evidence and Information Value

Weight of Evidence (WoE)

A technique that quantifies the predictive power of a categorical or binned continuous variable in relation to a binary target variable:

- Values are linear in the log-odds space
- Captures non-linear relationships via binning
- Handles categorical values

$$\mathrm{WoE}(X = a) = \log\left\{\frac{\mathbb{P}(X = a|\mathrm{Good})}{\mathbb{P}(X = a|\mathrm{Bad})}\right\}$$

Information Value (IV)

A technique based on the Kullback-Leibler Divergence, it measures the distance between probability distributions.

The higher the value the higher the predictive power of the variable.

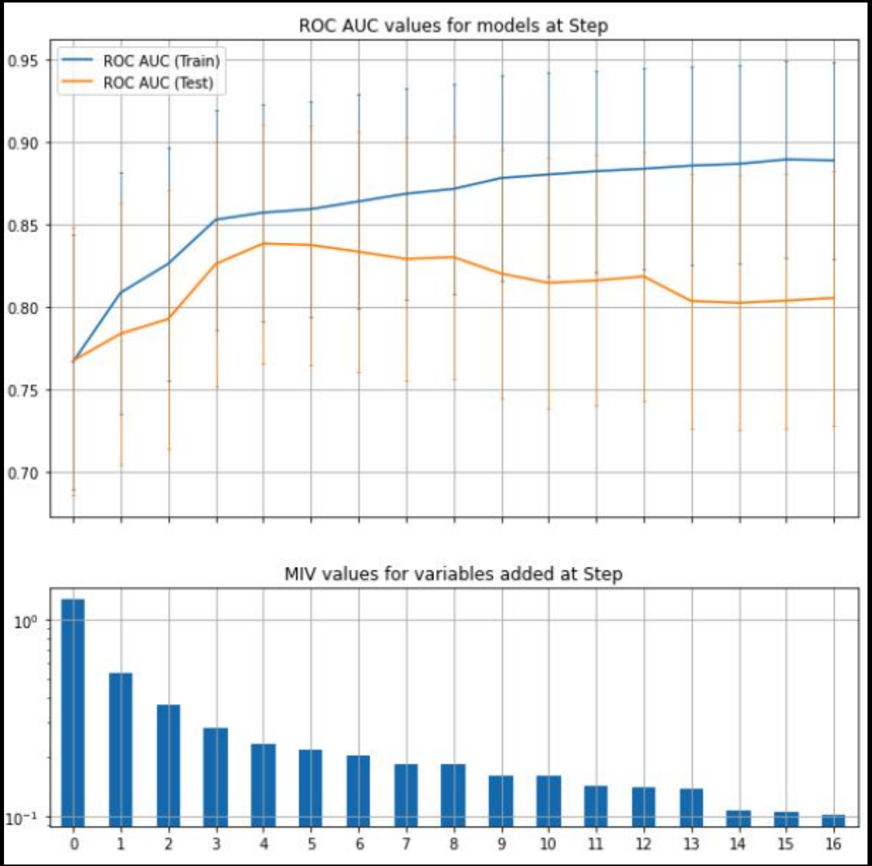
$$\mathrm{IV}(X) = \sum_{a\in X}\Big\{[\mathbb{P}(X = a|\mathrm{Bad}) - \mathbb{P}(X = a|\mathrm{Good})] \times \mathrm{WoE}_{\mathrm{observed}}(X = a)\Big\}$$

Marginal Information Value

Marginal Information Value (MIV)

An iterative feature selection technique that selects a parsimonious set of features that are largely independent of each other but collectively maximise model performance. It considers the information a new feature adds, given the information already captured by features already in the model.

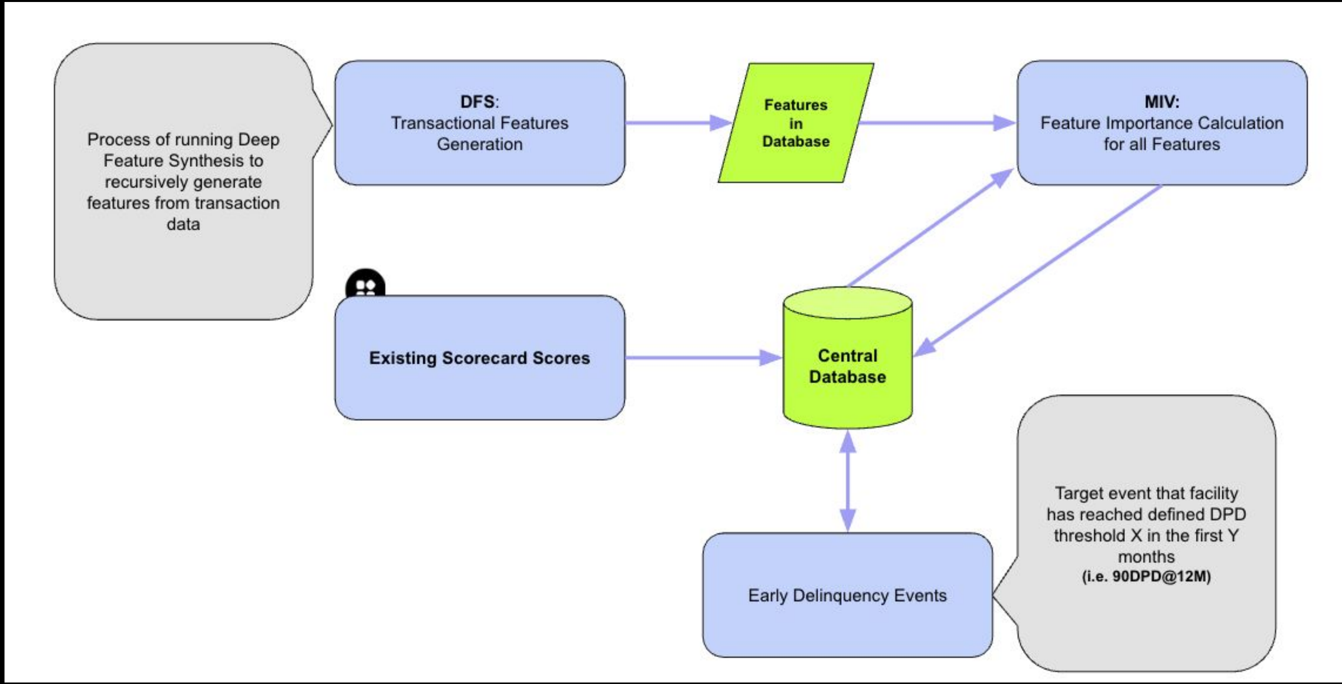
$$\mathrm{MIV}(X) = \sum_{a \in X} \Big\{ \big[\mathbb{P}(X = a|\mathrm{Bad}) - \mathbb{P}(X = a|\mathrm{Good})\big] \times \big[\mathrm{WoE}_{\mathrm{observed}}(X = a) - \mathrm{WoE}_{\mathrm{expected}}(X = a)\big] \Big\}$$



How DFS and MIV Combine Together

Background and Architecture

Schematic Presentation



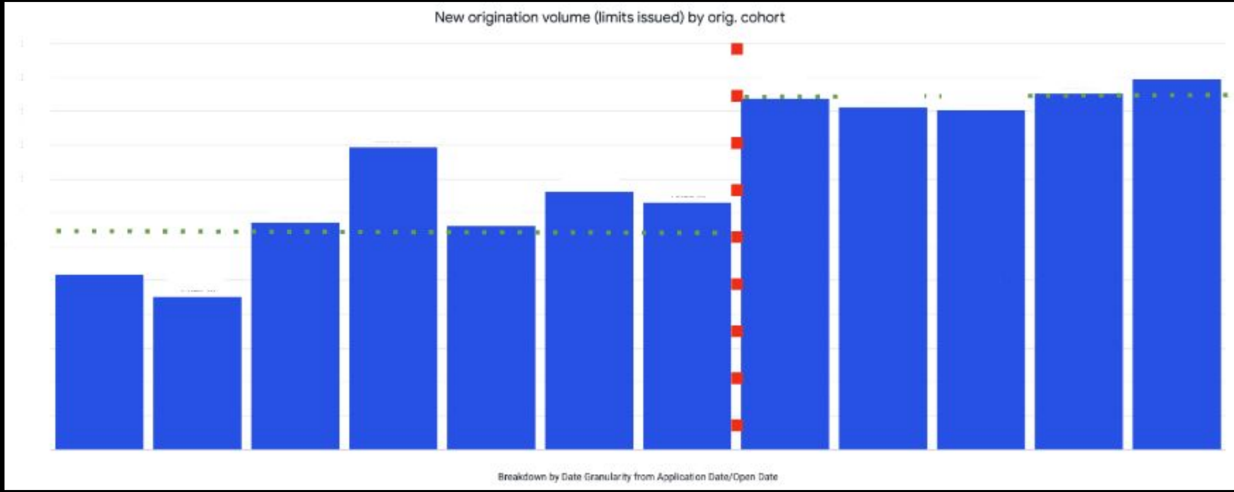
The process is recursive and features are added iteratively until model GINI on the test set stops increasing.

Business Benefits

Positive Selection

Case Study: Irish Personal Loans

- By utilising our own internal data, especially in markets where Revolut is a dominant player in the banking industry, or alternative data sources not commonly used in credit risk management, we can identify user segments that are **lower risk but under-served** in the market.
- Offering lower price than competitors, we will attract a larger share of these desirable low-risk, high-profitability users. These better users will choose our and not our competitor’s offers for credit, an effect called **positive selection.**
- When a model utilising internal Revolut transaction data was rolled out for personal loans, we saw **~30% increase in sales** driven primarily by higher offer take-up rate rather than increased approval rate, **at the same time reducing delinquency rates** of the portfolio.
- While usage of alternative data for credit is widespread in LatAm and Asia, we have the first mover advantage in many European countries.



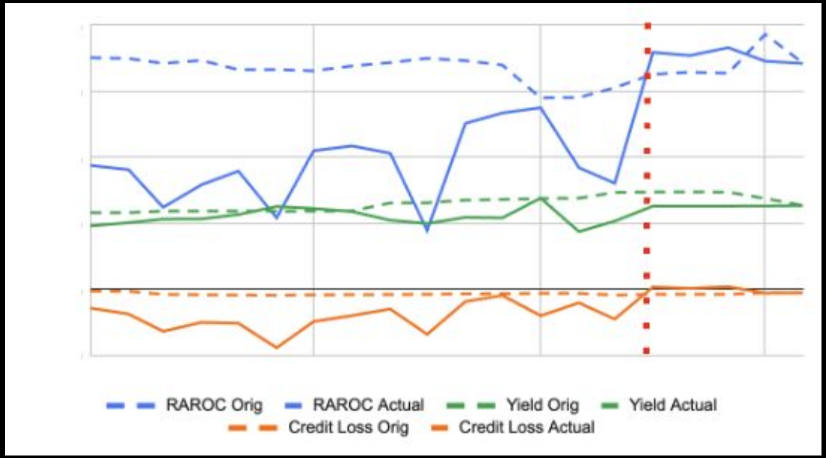
Incremental Disbursals *(each entry shows change in total loan amount issues in the category - we can see average loan amount increases for old accepts)*	**Rejected (Old Model)**	**Accepted (Old Model)**	**Overall**
Rejected (New Model)	0%	-8.9%	
Accepted (New Model)	+9.75%	+22.71%	
Overall			**+24%**

Business Benefits

Increased Risk-Adjusted Return on Capital

Case Study: Lithuanian Credit Cards

- By more accurately allocating customers to risk grades, we can price more appropriately for their risk. This increases RAROC, as shown by the data from a credit card portfolio.
- Identifying features that add predictive power to an acquisition PD model is identical to identifying user segments where we under- or over-predicted risk, as showcased in the previous sections. Therefore, identifying new data sources and predictive features allows to more accurately allocate a broader user base to risk grades, increasing RAROC of the credit originations.



Future Research

- Explore the use of LLMs to generate even more nuanced features based on the relationship between data sources
- Leverage the use of embeddings as model inputs to capture even more nuances in the data

References

- The Automatic Statistician (link)
- Deep Feature Synthesis, Towards Automating Data Science Endeavours (link)
- Automated Feature Generation from Structured Knowledge (link)
- Intelligent Credit Scoring, Naeem Siddiqi
- Credit Risk Scorecards: Development and Implementation Using SAS, Mamdouh Refaat

Thank you