

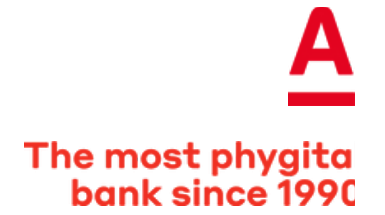
Siberian Alfa Hack

A
The most phygita
bank since 1990



Команда ES

Задача и датасет



Задача: Предсказать отток клиентов на предоставленном организаторами датасете

Датасет:

100 параметров, содержащих информацию о клиентах и их активности внутри банка.

Из них 9 категориальных параметров, остальные числовые.

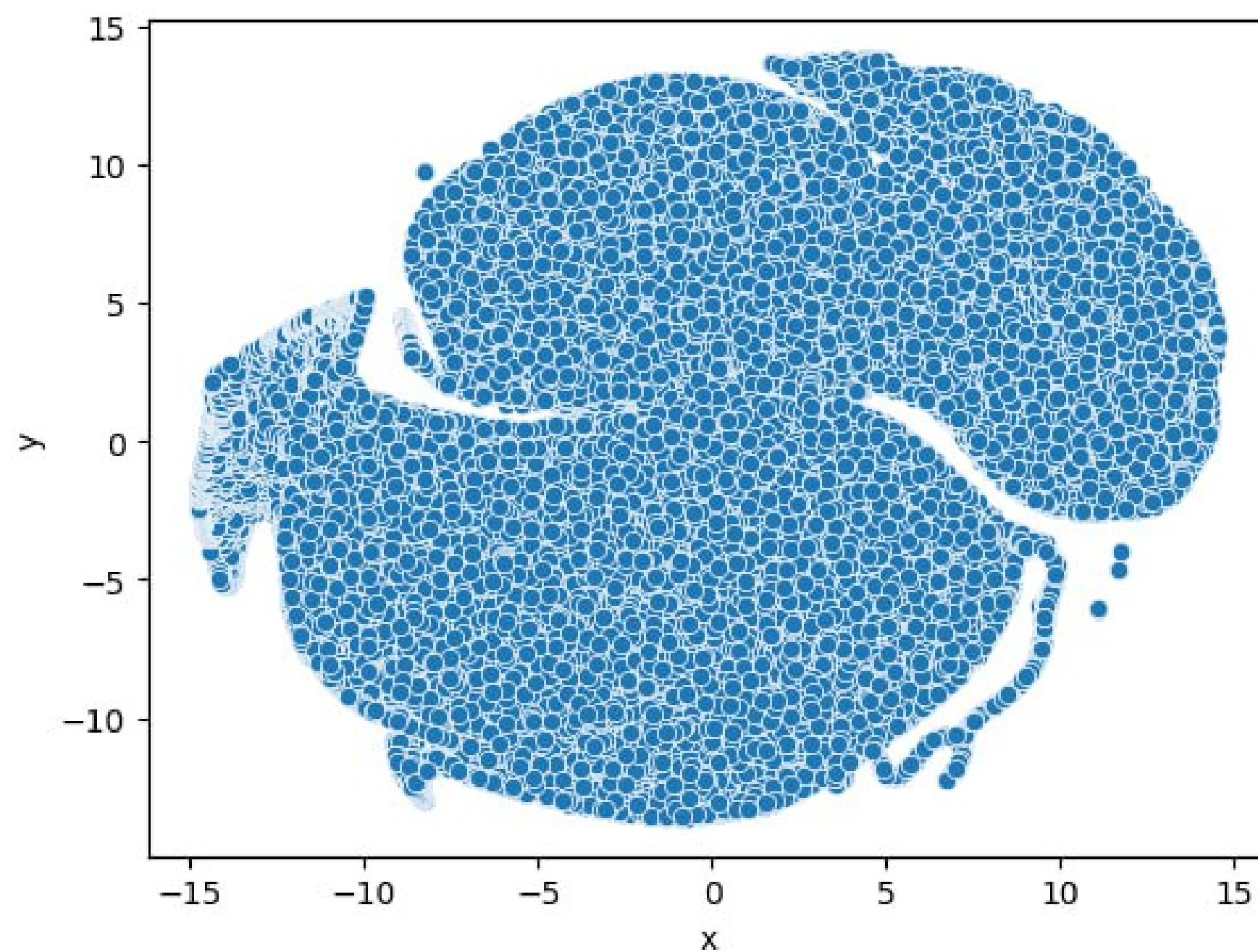
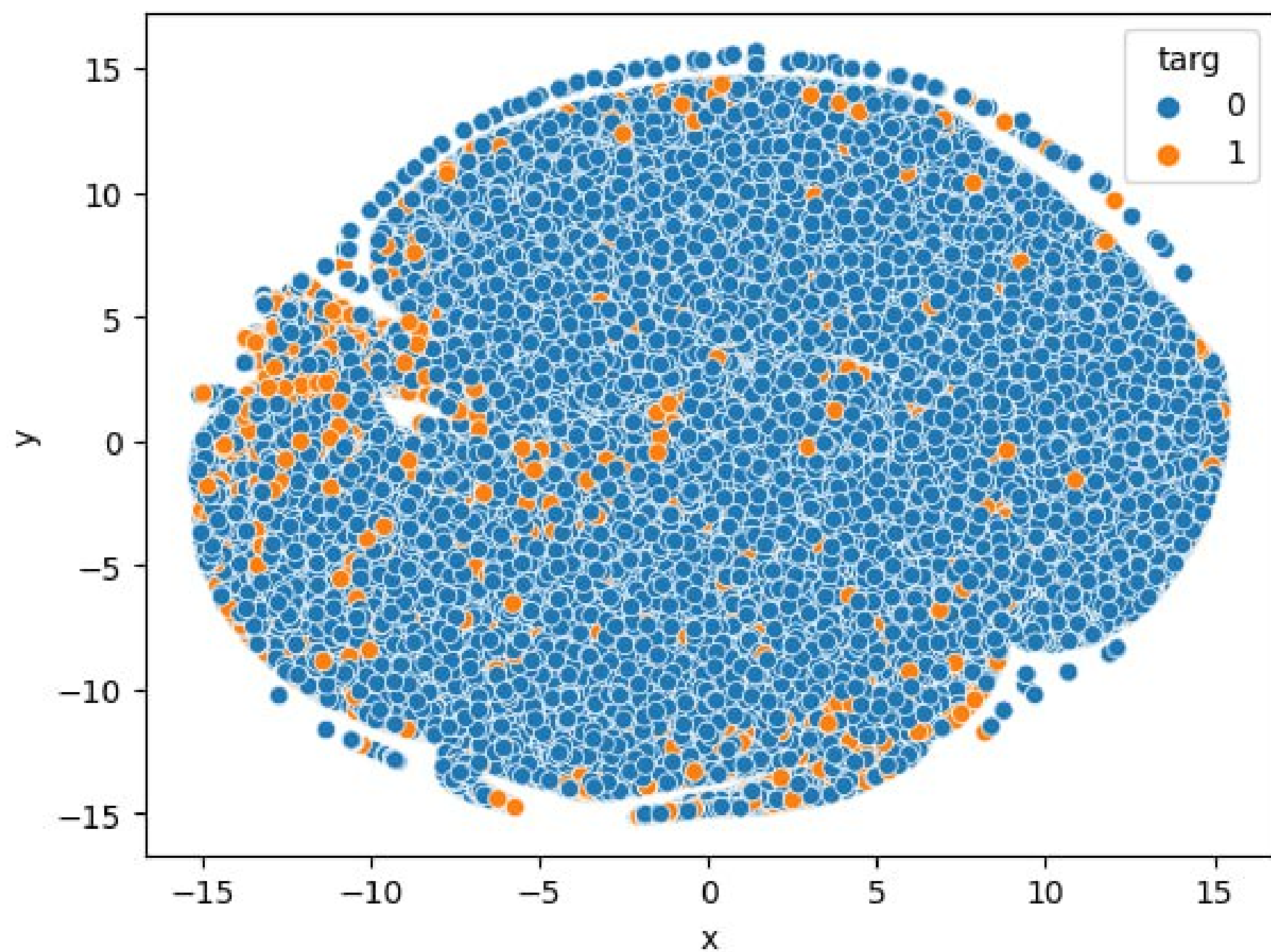
Train состоит из 360 тыс. строк, test из 100 тыс.

Предобработка данных

Основные этапы предобработки:

- Удалены логические ошибки в данных: количество дней в характеристиках < 0
- Обработаны значения -999 и None в pr.nap в датасете
- Удалили колонки, где количество NaN $> 60\%$
- Разбили ОКВЕД на группы
- Нагенерировали трендов с помощью временных колонок
- Нагенерировали индикаторных признаков

T-SNE



Решение

Основное решение было разработано на основе LightGBM с параметрами, подобранными с помощью Optuna, но были опробованы такие подходы как:

- Логистическая регрессия
- Catboost, XGBoost
- Random Forest
- Наивный Байес



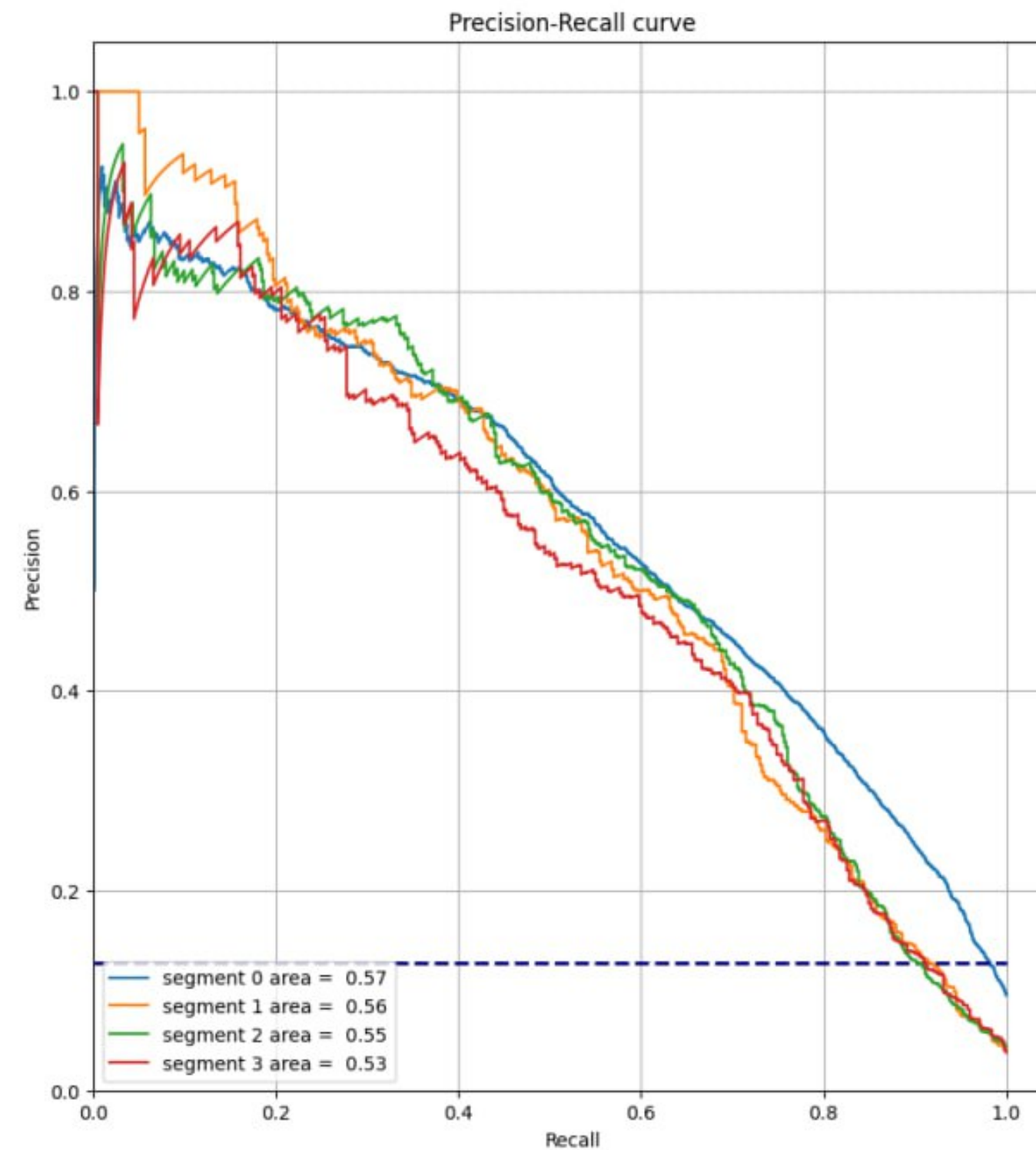
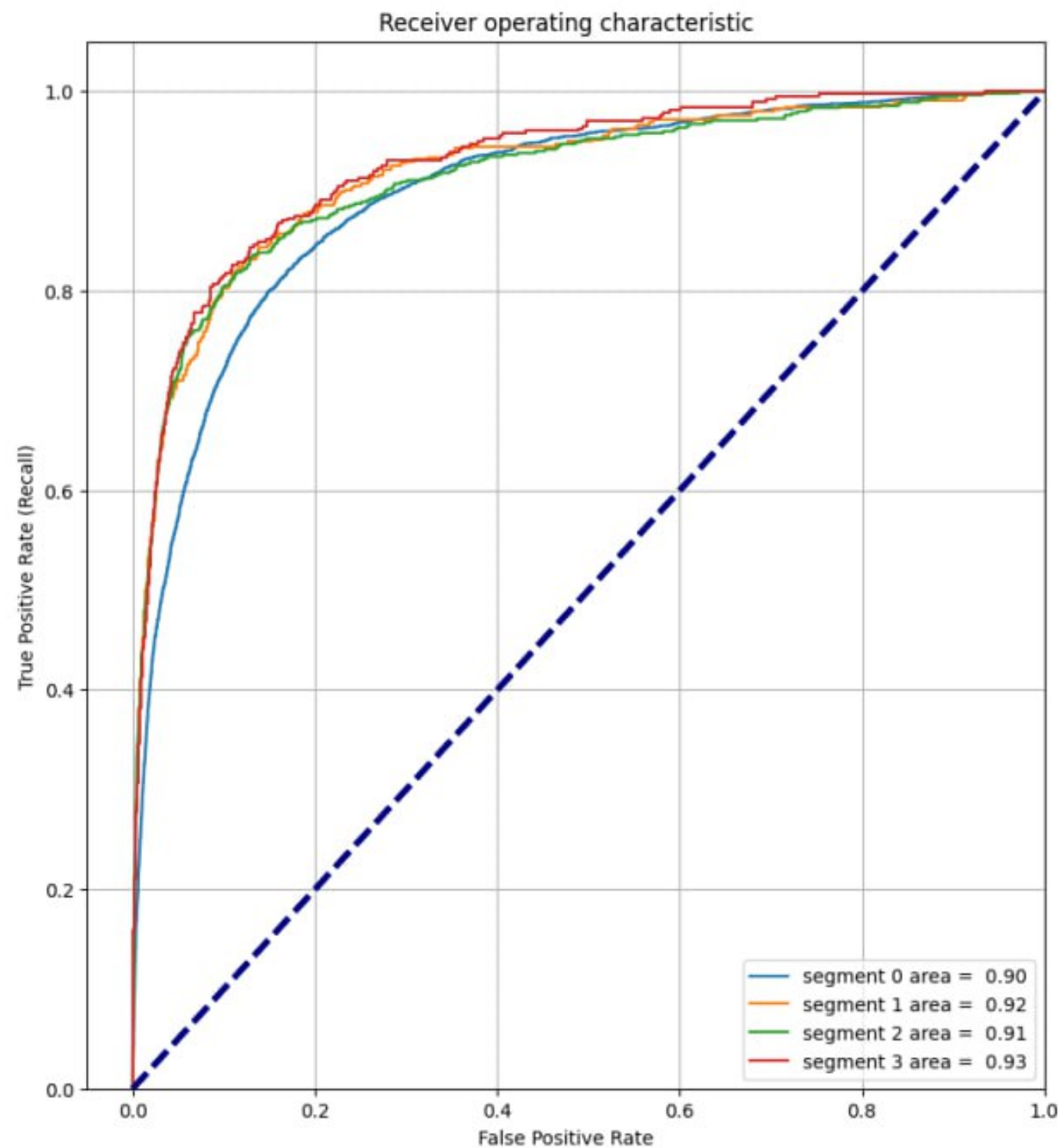
На тестовом датасете

- ROC-AUC: 0.92
- RECALL: 0.33

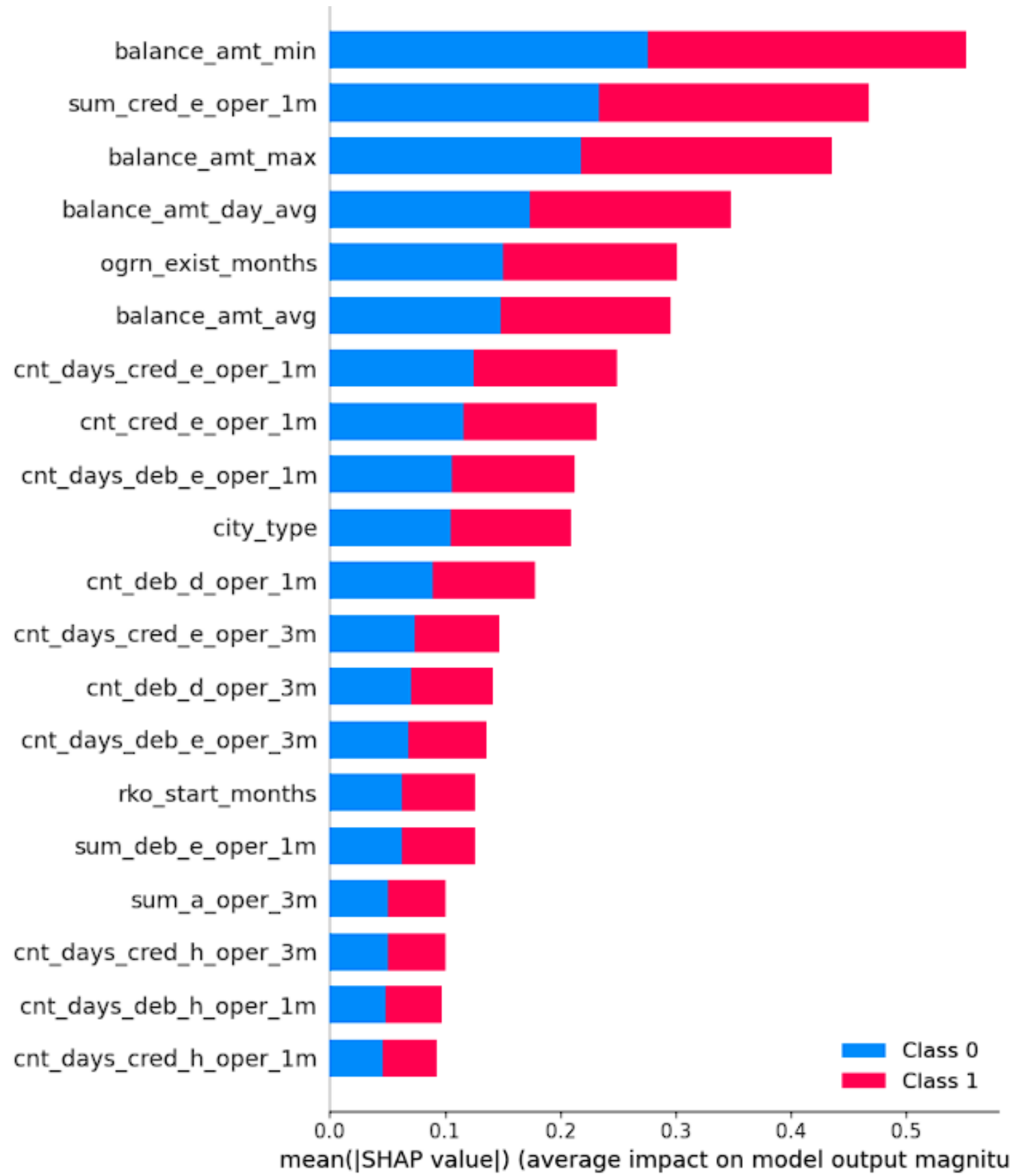
ROC-AUC по различным сегментам



The most phygita
bank since 1990



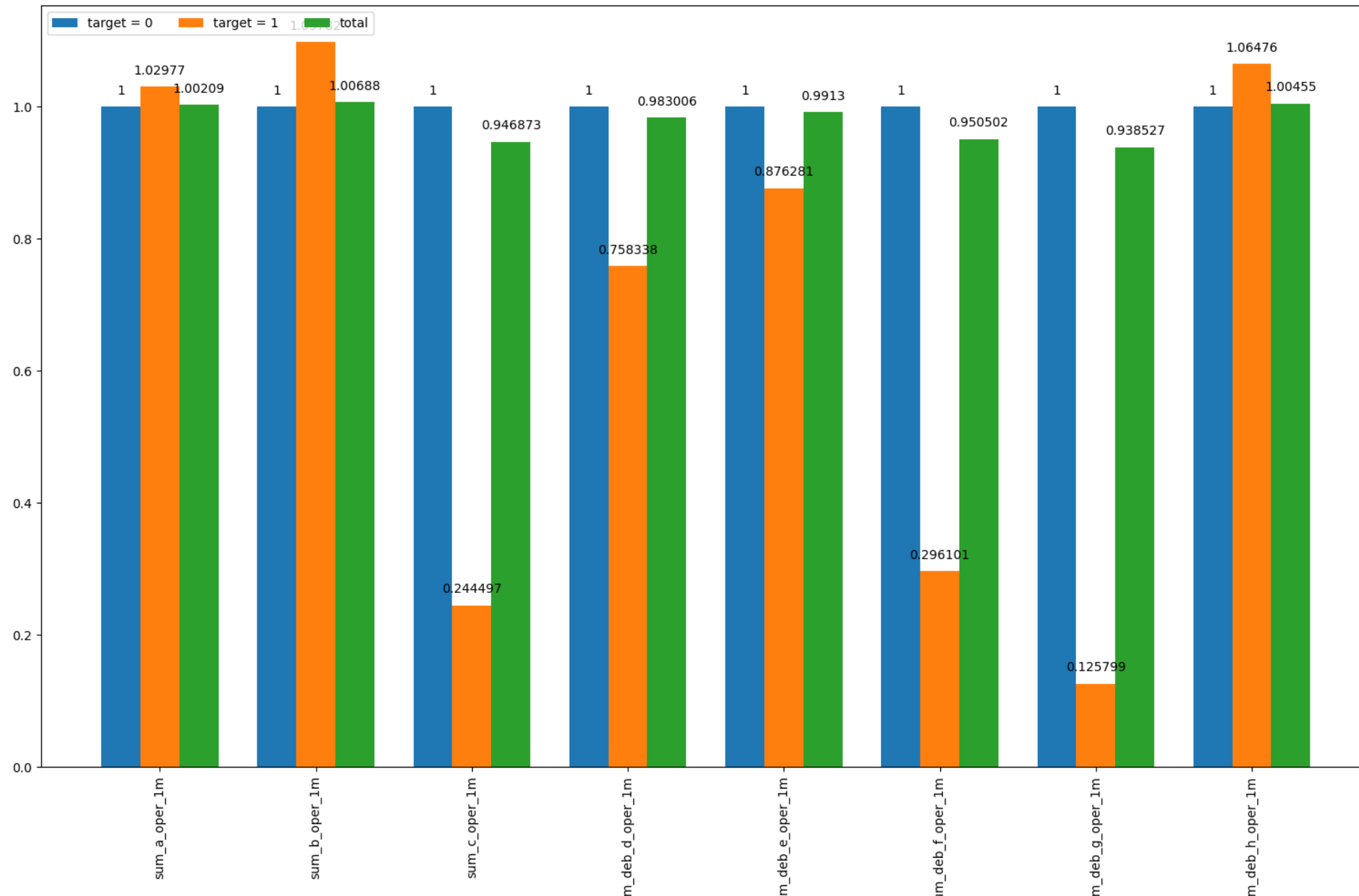
Влияние признаков



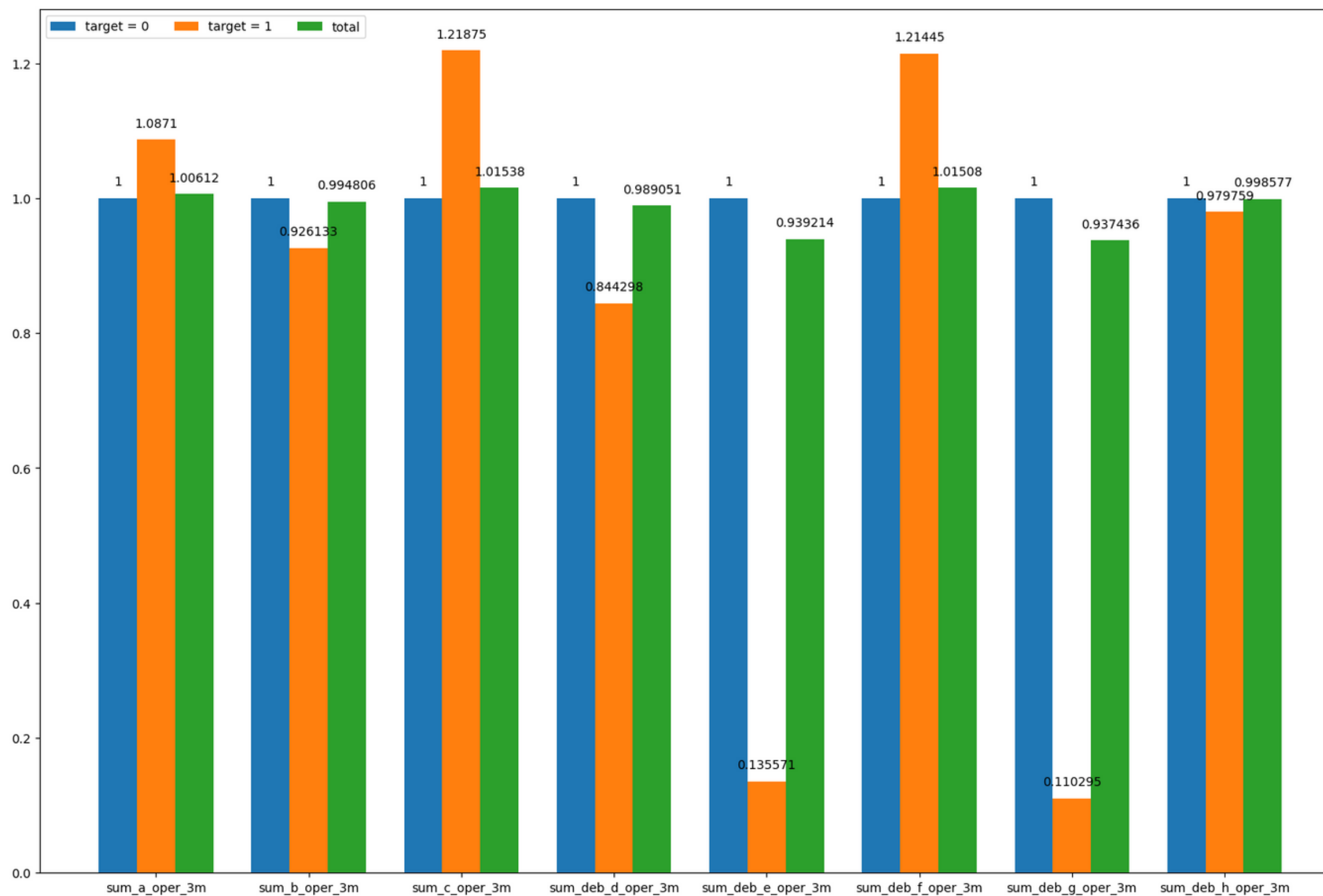
Как предсказать отток без ML



The most phygita
bank since 1990



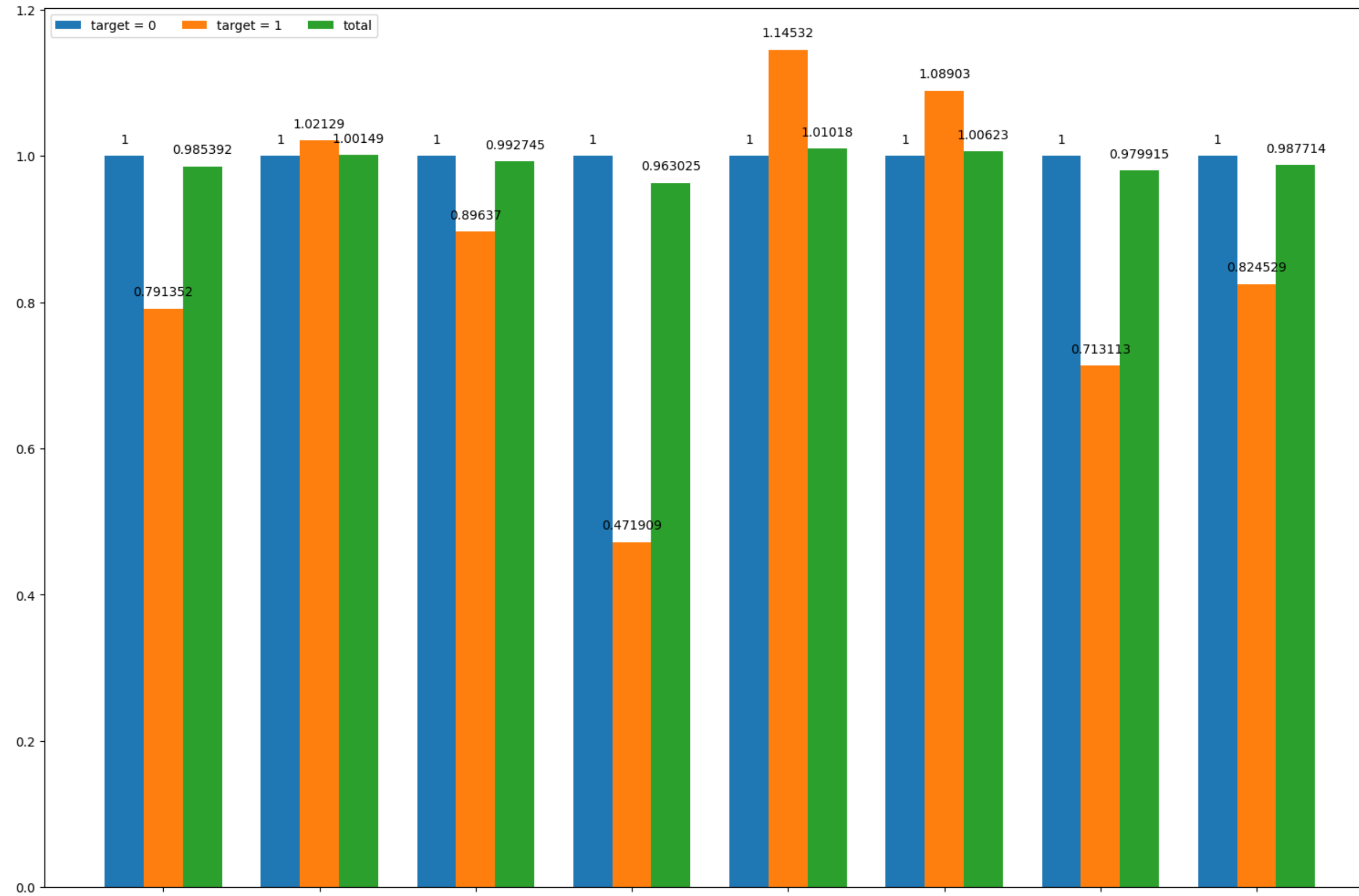
Как предсказать ОТТОК без ML



Как предсказать отток без ML



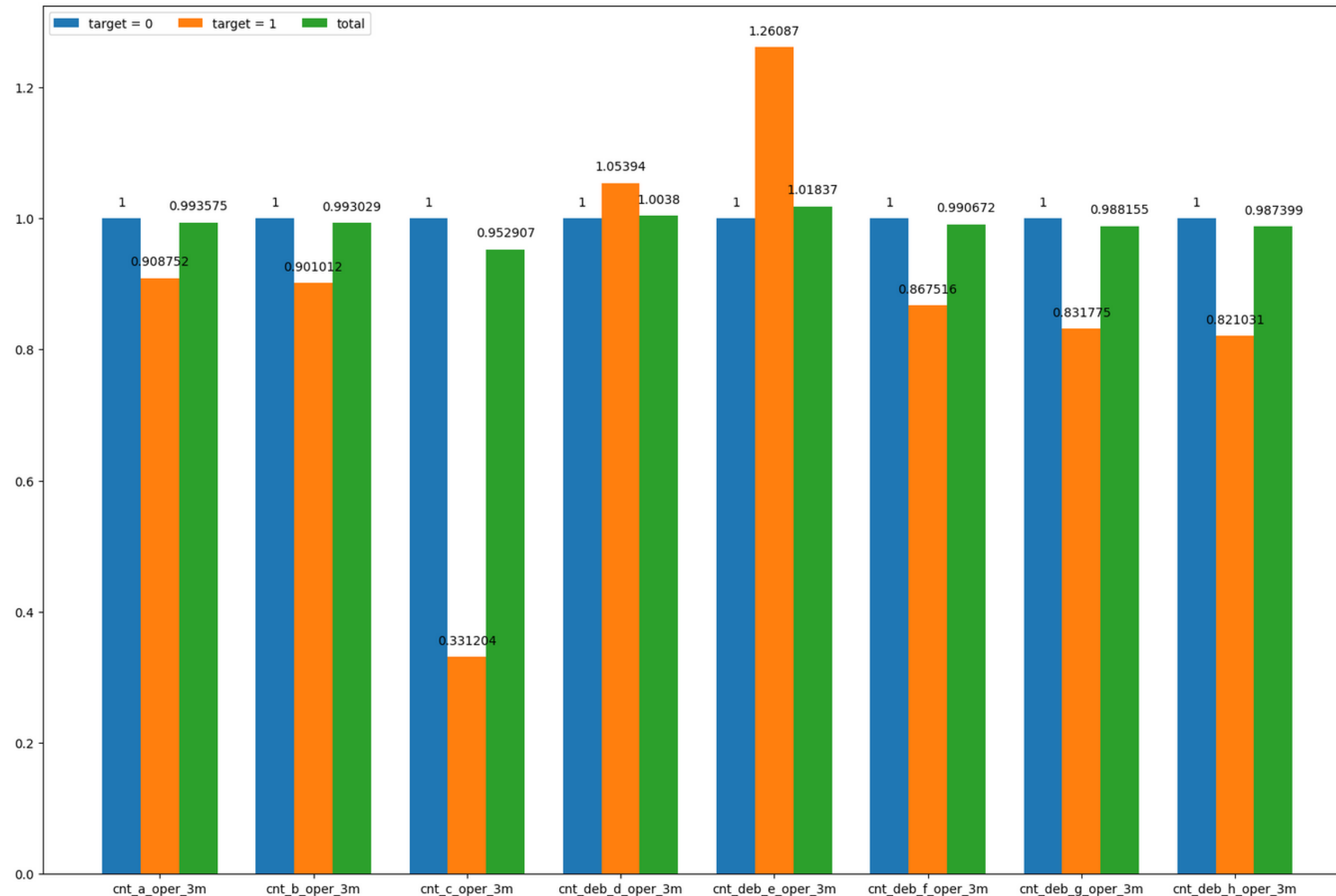
The most phygita
bank since 1990



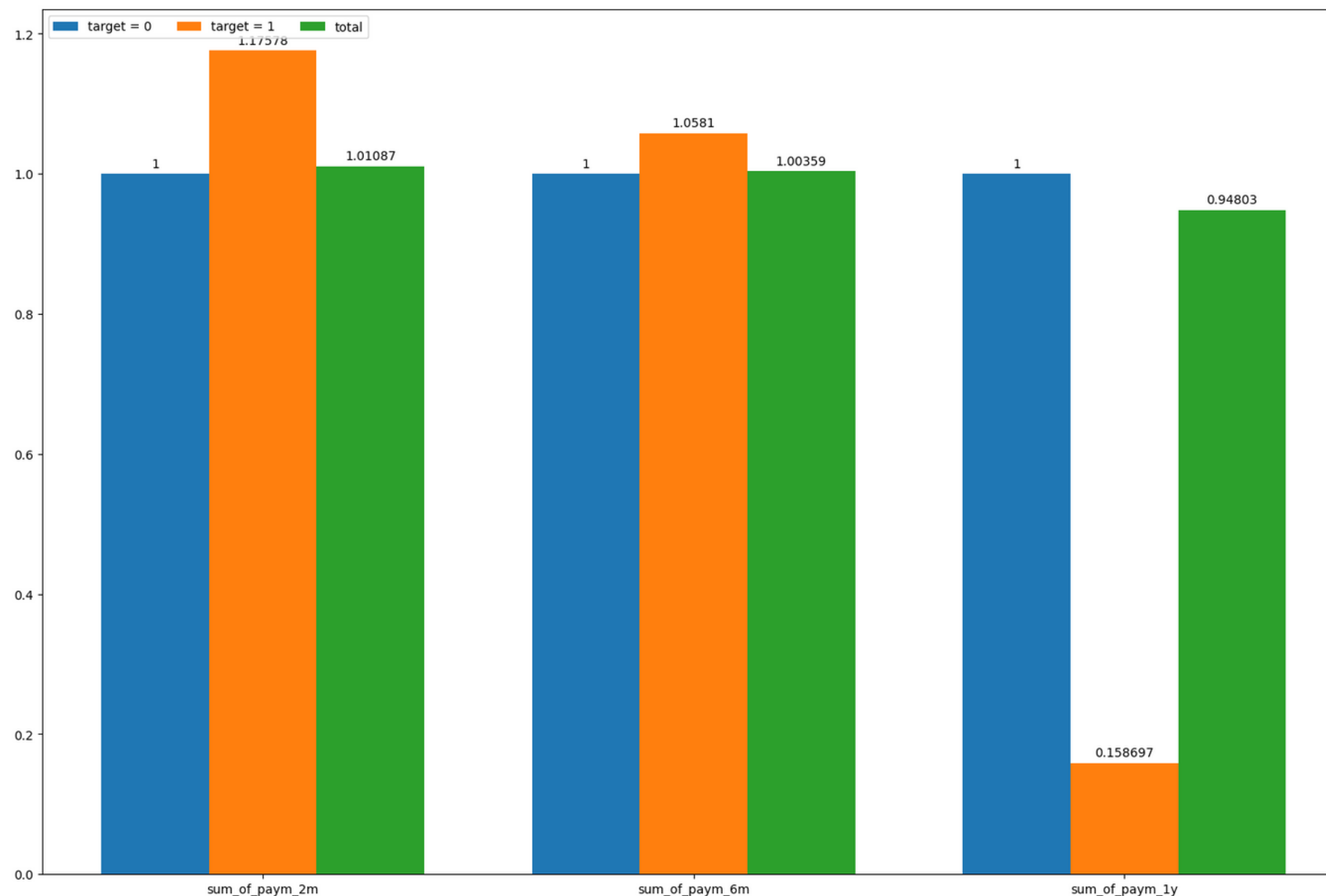
Как предсказать отток без ML



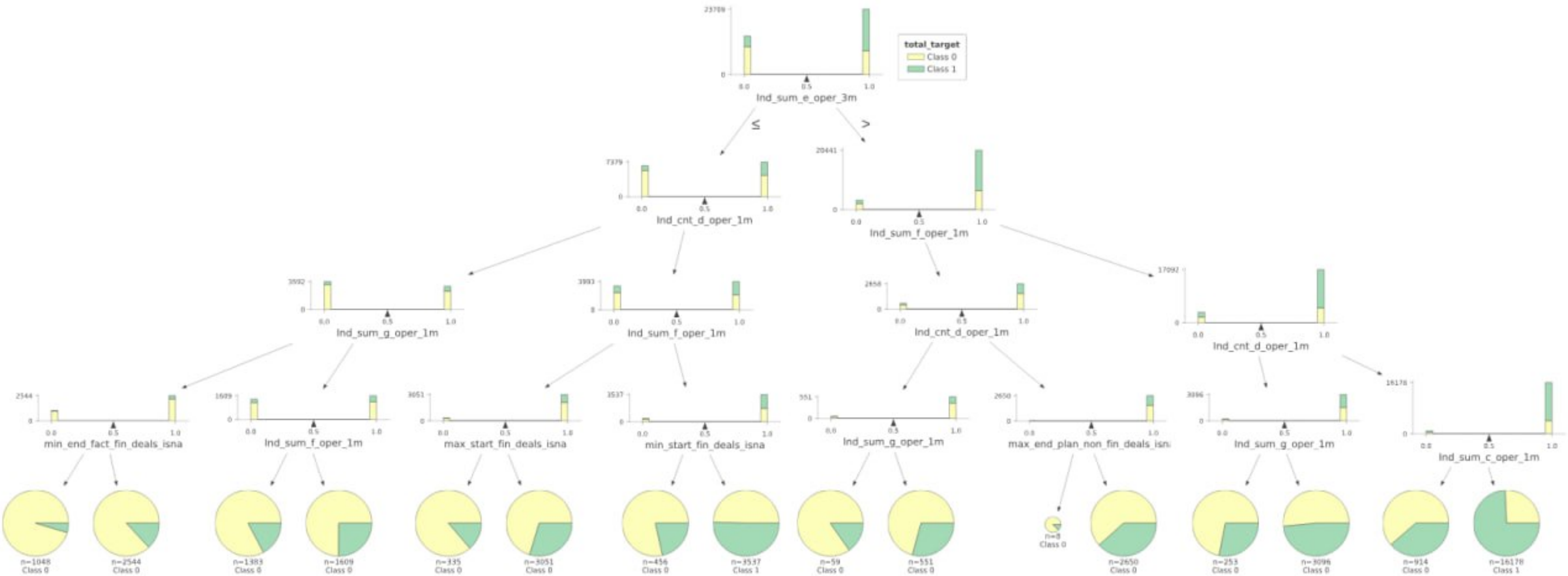
The most phygita
bank since 1990



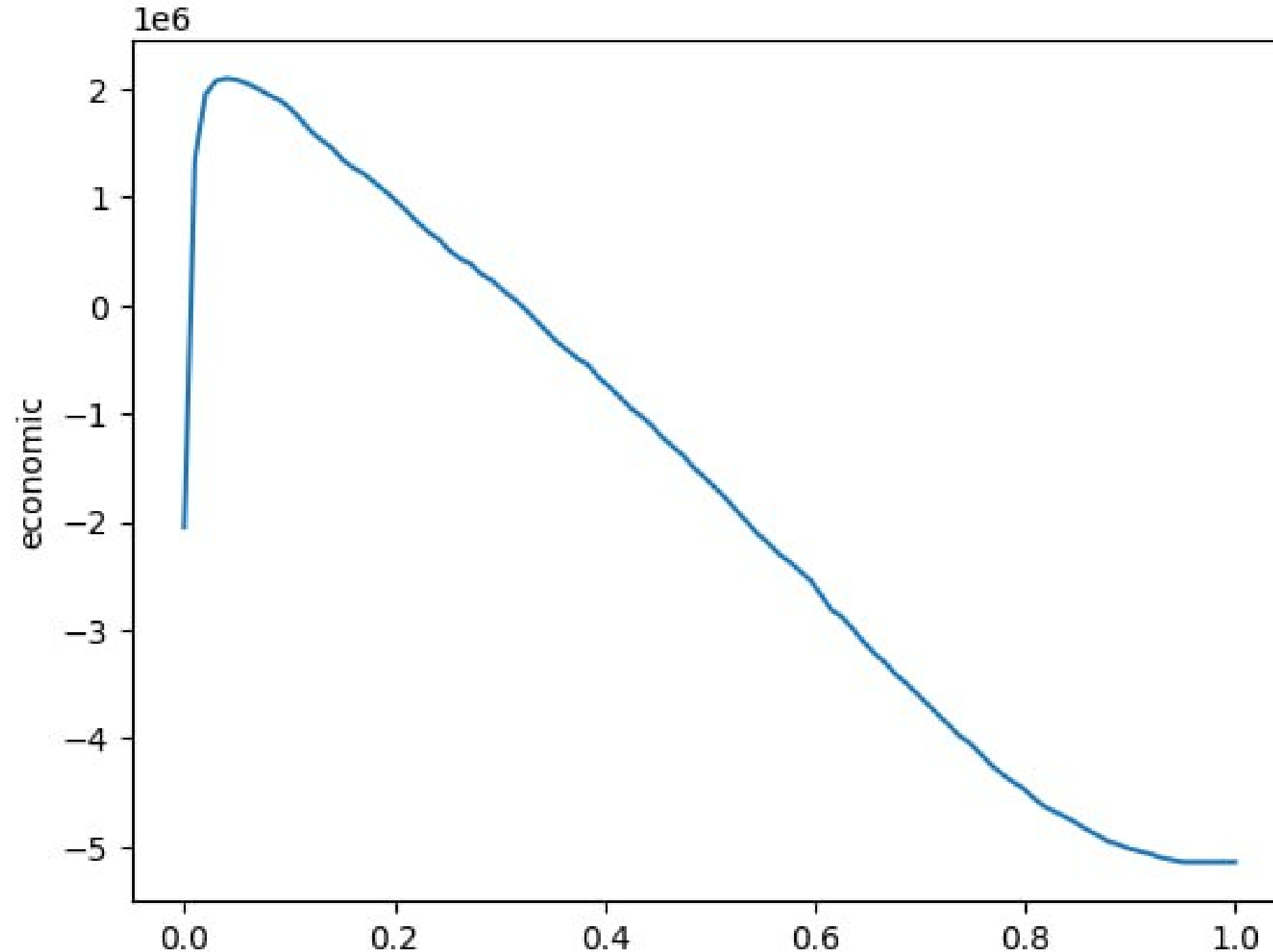
Как предсказать отток без ML



Интерпретируемая модель



Зависимость размера прибыли от разных порогов модели



Считаем, что удержать
клиента стоит
100 условных
единиц, а
потерять клиента - 1000

Выводы

Одними из основных показателей оттока является снижение финансовой активности, в частности операции типа C, E, D, F, показатели баланса.

Наиболее важными для модели являются признаки:

1. Минимальный остаток на счетах клиента за месяц
2. Сумма исходящих операций "E" за 1 месяц
3. Максимальный остаток на счетах клиента за месяц