

Проект SQL: Анализ оттока клиентов банка

Code ▾

Изгарова Анна, aiizgarova

Задача

Отток клиентов банка. Требуется определить, с чем связано то, что клиенты не заключают новые договоры. Также необходимо предложить вариант - как сократить уход клиентов. Удержания клиентов, повышения их уровня лояльности важно для сбережения ресурсов на привлечение новых клиентов.

Анализ

Данные и логика анализа

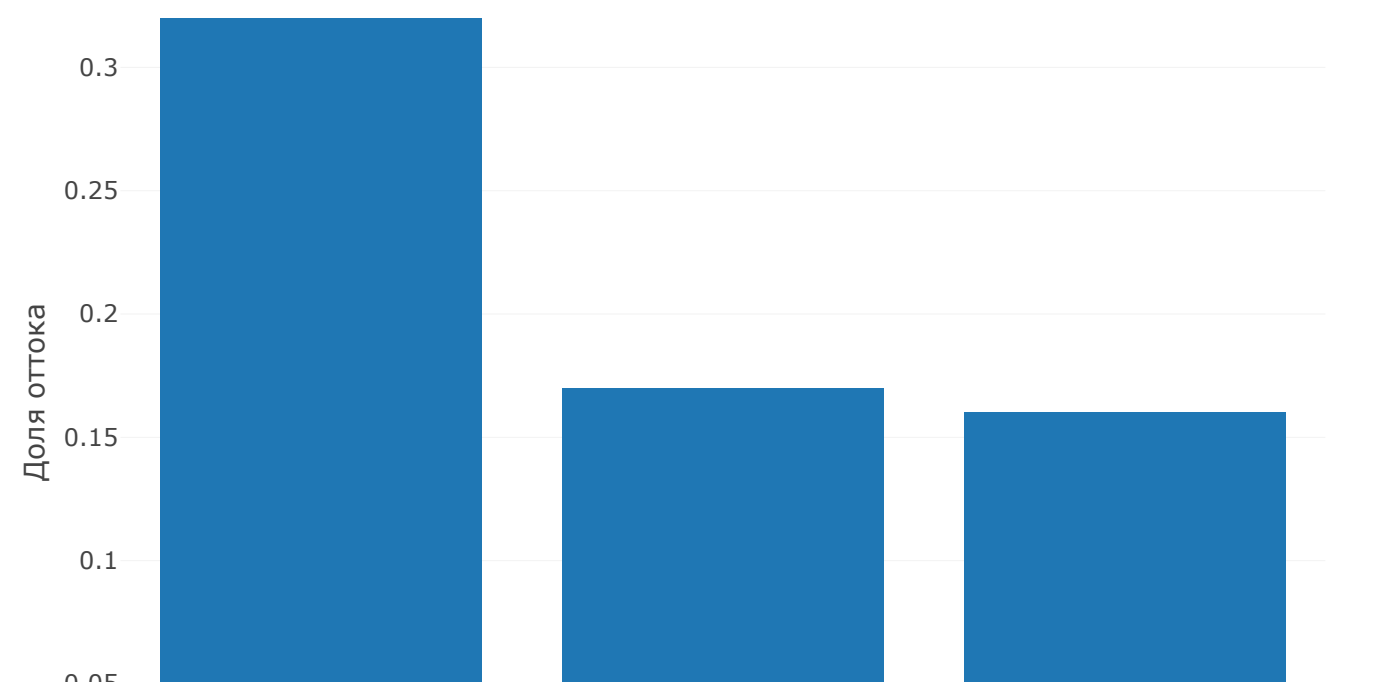
В отчете используются данные из базы данных MariaDB → dbname='bank'. Подключение к ним происходит при помощи SQL.

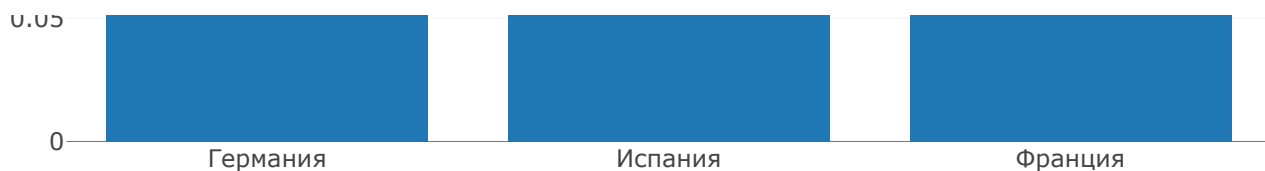
Используя SQL-запрос, посмотрим на пропорцию оттока клиентов:

Уход клиента	Доля от общего количества
1	0.2037
0	0.7963

Около 20% клиентов уходят.
Можно предположить, что этот показатель будет различаться в зависимости от географии. Тогда можно рассмотреть задачу сокращения оттока клиентов по конкретной стране, чтобы найти какой-то наиболее подверженный риску филиал банка.
Рассмотрим распределение, полученное из данных так же при помощи SQL-запроса:

Отток клиентов в зависимости от страны





В Германии клиенты банка уходят чаще. Для сужения анализа следует выбрать группу клиентов, чья страна указана как “Германия”.

Модель

Для предсказания оттока строятся две модели. Их точности сравниваются и выбирается лучшая модель из двух. Из базы данных выгружаются выборка резидентов Германии, выбираются только необходимые для анализа переменные. Данные разбиваются на тестовую и обучающую выборки. Что касается технической подготовки данных: устраняются скоррелированные количественные переменные, категориальные преобразуются в дамми-переменные, данные нормируются, некоторые - логарифмируются.

Для оценки качества моделей используются метрики: sensitivity, specificity, roc_auc.

Логистическая регрессия

Метрика	Estimator	Estimate
sens	binary	0.9037328
спес	binary	0.3551020
roc_auc	binary	0.7343571

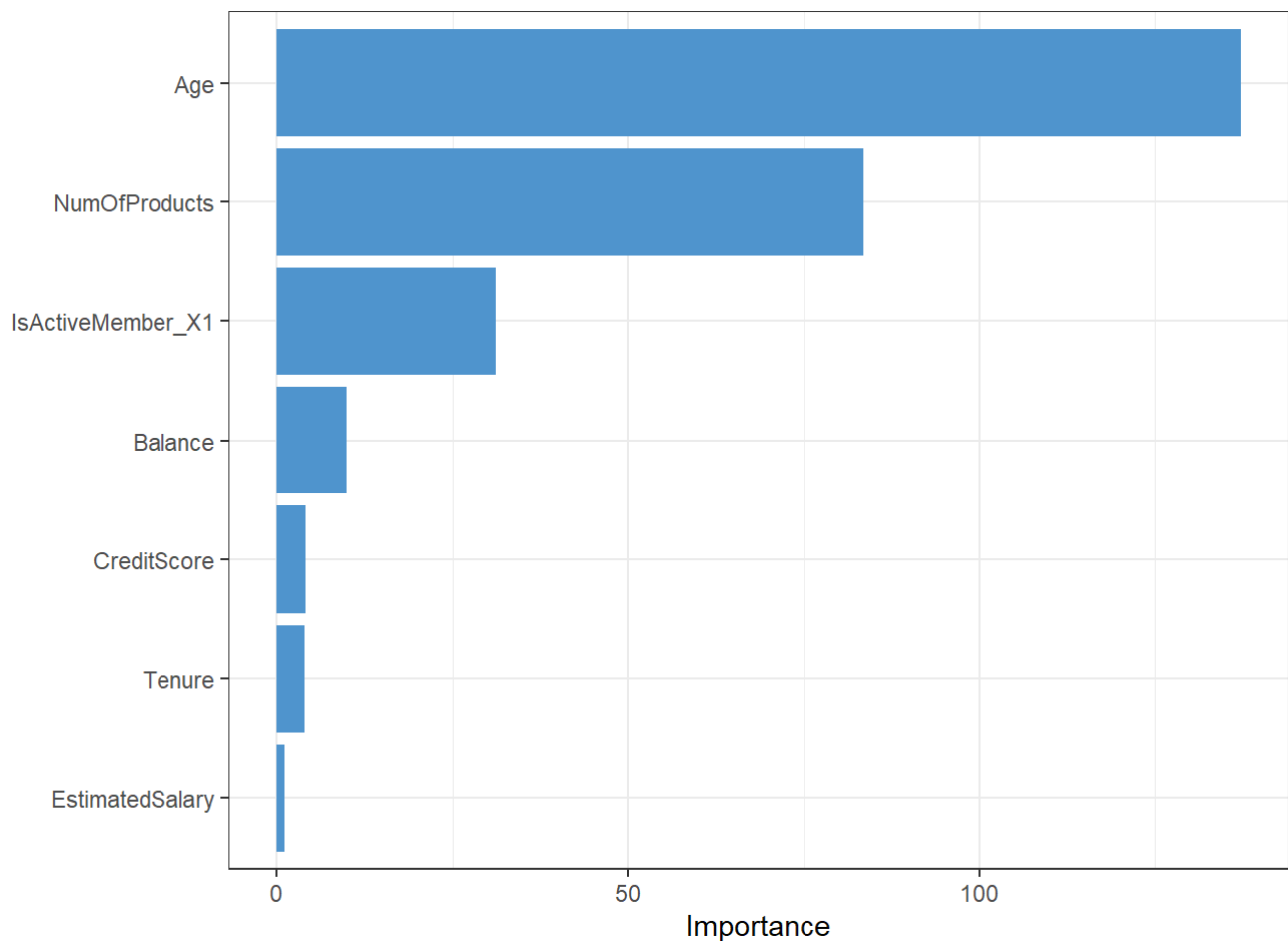
Дерево

Метрика	Estimator	Estimate
sens	binary	0.9214145
спес	binary	0.5591837
roc_auc	binary	0.7714125

Судя по ROC_AUC дерево в нашем случае предсказывает отток лучше логистической регрессии. Выбираем дерево.

Симуляция

Следует предложить меры сокращения оттока, а также смоделировать их внедрение, чтобы выяснить, становятся ли клиенты более лояльными. Для этого произведем оценку важности каждой переменной, влияющей на итог:



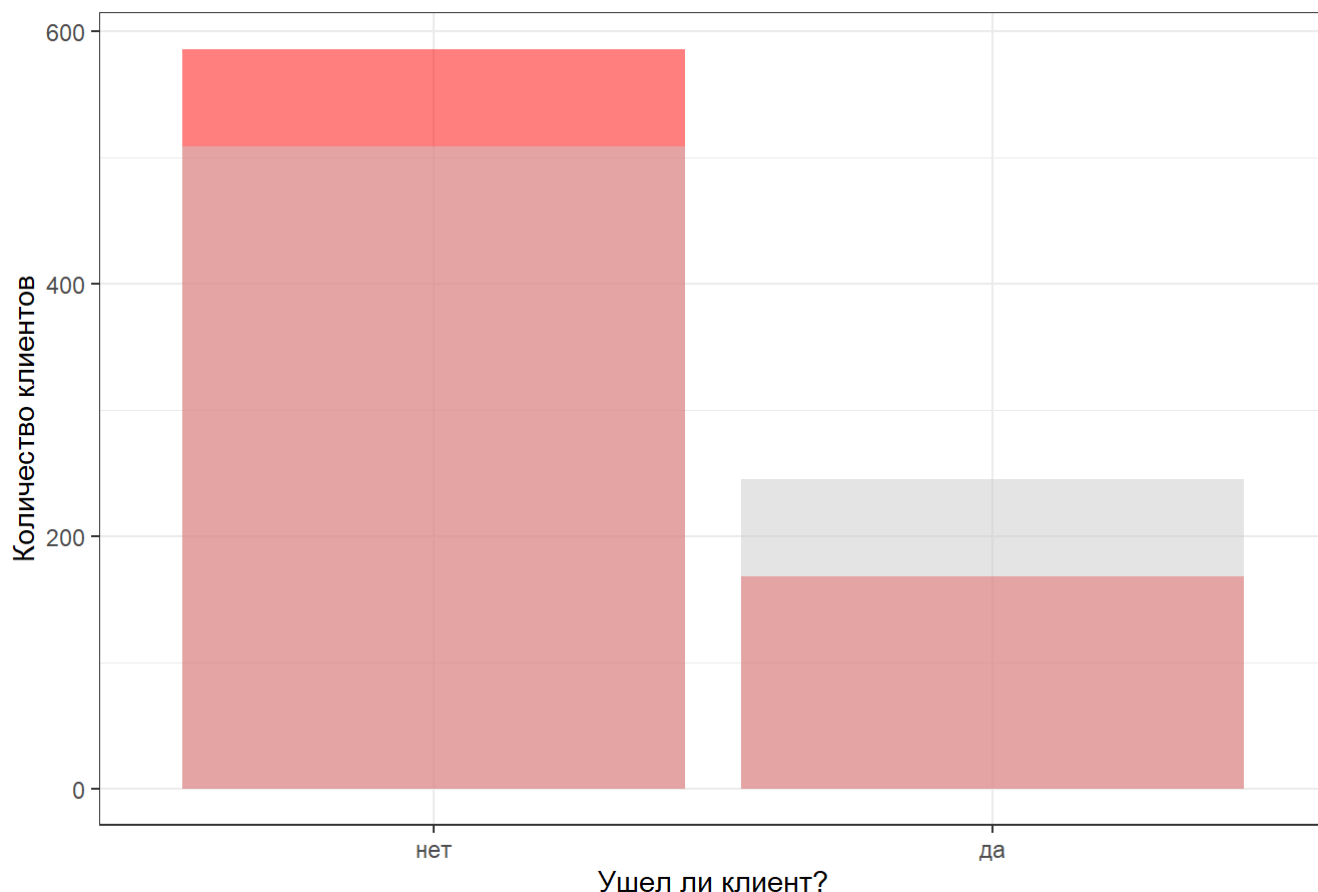
Из графика видно, что важным оказалось количество продуктов банка, используемых клиентом (2-я строка).

Предложение:

Рекомендовать клиентам новые продукты - повысить количество продуктов, которыми пользуется клиент (кредитов, карт, счетов). Например, заключить страховку, открыть инвестиционный счет, сберегательный счет или подключиться к мобильному оператору банка. Это должно повысить лояльность клиента и увеличить издержки клиента на выход из экосистемы банка.

Поэкспериментируем. Попробуем предложить 1 продукт банка клиентам, которые сейчас пользуются только одним продуктом банка. Считаем, что с вероятностью 7% клиенты соглашаются пользоваться этим +1 продуктом.

Изменение в оттоке клиентов после нововведения



Новое значение изображено красным, прежнее - серым. Как видно, отток клиентов в Германии сокращается, если предлагать им дополнительный продукт банка.

Общие выводы

Цель работы заключалась в предсказании оттока клиентов банка. В ходе работы была выбрана группа клиентов, относящихся территориально к Германии. На основании данных о точности двух моделей (логистическая регрессия и дерево) было принято решение использовать дерево в качестве предсказательной модели. Выяснилось, что значимым сигналом оттока клиентов является низкое количество продуктов банка, которыми пользуется клиент. Поэтому было выдвинуто предложение рекомендовать клиентам, которые пользуются только 1 продуктом, еще +1 дополнительно. Была смоделирована ситуация, когда клиенты соглашались в 7% случаев. Такое нововведение дало положительную динамику: клиенты стали уходить меньше. Таким образом, был обнаружен действенный способ снизить отток клиентов - предлагать им дополнительные продукты банка. Это повысит лояльность и удержит клиентов.