

Альфа Банк



Задача прогнозирования баланса клиентов

Предскажите, какие клиенты банка сохранят положительный баланс на своих счетах, основываясь на данных о транзакциях и характеристиках клиентов. Вам необходимо разработать модель, которая классифицирует клиентов по минимальному балансу в следующем месяце.



Постановка задачи

Разработайте модель для прогнозирования минимального баланса клиентов банка в следующем месяце. Модель должна классифицировать клиентов по их минимальному балансу, оценивая вероятность попадания в одну из 7 категорий.

Для этого:

- 01** Изучите предоставленные данные о клиентах и их транзакциях за трехмесячный период.
- 02** Проанализируйте взаимосвязи между признаками и целевой переменной.
- 03** Постройте модель на тренировочных данных и оцените ее качество.
- 04** Сделайте предсказание для тестовой выборки и сохраните его в формате CSV.
- 05** Отправьте решение на платформу в виде файла предсказания и узнайте качество предсказания в виде оценки с помощью метрики WMAE.
- 06** Доработайте модель и отправьте улучшенные предсказания для повышения результата.



ВАЖНО!

Вы имеете право загружать ваше предсказание на платформу в формате CSV не более 3 раз в течение одного дня (в промежутке между 00:00 и 23:59 мск).

Бизнес-контекст

Прогнозирование минимального баланса клиентов банка — это определение самого низкого остатка средств на счете клиента в следующем месяце.

Для решения этой задачи банк классифицирует клиентов по 8 категориям:

- Класс 0 — отсутствие средств на счете.
- Классы с 1 по 6 — различные уровни минимального баланса (где большее значение класса указывает на больший остаток, т. е. остаток у клиентов класса 1 меньше, чем у класса 2, а класс 6 имеет самый высокий минимальный остаток на счете).

Если клиент попадает в класс 0, банк разрабатывает стратегию по удержанию клиента через программы лояльности, персональные предложения или изменение условий обслуживания.

ДЛЯ КАКИХ ЗАДАЧ ИСПОЛЬЗУЕТСЯ ПРОГНОЗИРОВАНИЕ БАЛАНСА

1. Оптимизация кредитного лимита

Банку необходимо точно понимать уровень дохода клиента, чтобы предложить ему оптимальный кредитный лимит:

- Слишком низкий лимит снижает удобство использования банковских продуктов.
- Слишком высокий лимит увеличивает риск выхода клиента в дефолт.

Прогноз минимального баланса помогает находить баланс между рисками и возможностями.

2. Контроль кредитной нагрузки

Соблюдение требований Центрального банка по предельной долговой нагрузке (ПДН) позволяет:

- Избегать санкций в сторону клиента и снижать вероятность просрочки по платежам.
- Предлагать клиентам адекватные финансовые продукты, соответствующие их платежеспособности и клиентскому профилю.

3. Таргетированные предложения

Благодаря анализу минимального баланса банк может проводить сегментацию клиентов по их платежеспособности:

- Клиентам с высоким балансом предлагать премиальные услуги (инвестиционные счета, VIP-обслуживание, премиальные банковские карты и уникальные условия).
- Клиентам с низким балансом — льготные кредитные предложения, кешбэк, бонусные программы.

4. Управление ликвидностью

Банк заранее знает, сколько денег останется на счетах клиентов, что позволяет:

- Эффективно планировать кредитные и инвестиционные программы.
- Оптимизировать платежные процессы и управление ликвидностью.



Критерии успеха

Основным критерием успешной модели считается ее способность точно предсказывать минимальный баланс клиентов в следующем месяце. Для оценки качества модели используется метрика WMAE, которая позволяет учитывать вес конкретного класса, предсказание которого наиболее финансово выгодно для банка при работе с конкретным клиентом. Предсказания выполняются на тестовых данных, что обеспечивает независимую оценку качества модели и позволяет выявить возможное недо- или переобучение.

Ограничения

Для создания и настройки модели вы можете использовать язык Python версии 3.10 и выше. Также вы можете использовать любую библиотеку для машинного обучения и предобработки данных, являющуюся Open-Source-ресурсом. Использование закрытых библиотек, частных API и чужого кода, не подпадающего под разрешение о свободном распространении и использовании, запрещено.

Источники информации

(Ссылка на данные на Яндекс Диск)

Полный датасет был разделен на несколько датасетов одинаковой структуры и разного содержания.

- Тренировочный датасет **train_transactions.parquet**.
- Тестовый датасет **test_transactions.parquet**.
- Каждый из датасетов содержит следующие признаки:
 - **client_num** — уникальный анонимный идентификатор клиента;
 - **date_time** — дата и время совершения транзакции;
 - **mcc_code** — MCC-код торговой точки;
 - **merchant_name** — закодированное имя мерчанта (продавца);
 - **amount** — сумма проведенной транзакции.
- Файл **train_target.csv** — файл с конечными классами клиентов, информация о которых описана в тренировочных данных. Содержит признаки:
 - **client_num** — уникальный анонимный идентификатор клиента;
 - **target** — целевая переменная (метка класса клиента от 0 до 6).
- Файл **test_target.csv** — файл о клиентах, значение целевой переменной для которых только предстоит определить. Содержит признаки:
 - **client_num** — уникальный анонимный идентификатор клиента.
- Файл **sample_submission.csv** — пример сабмита для отправки на платформу. Содержит признаки:
 - **client_num** — уникальный анонимный идентификатор клиента;
 - **target** — вероятность перехода клиента в определенный класс.

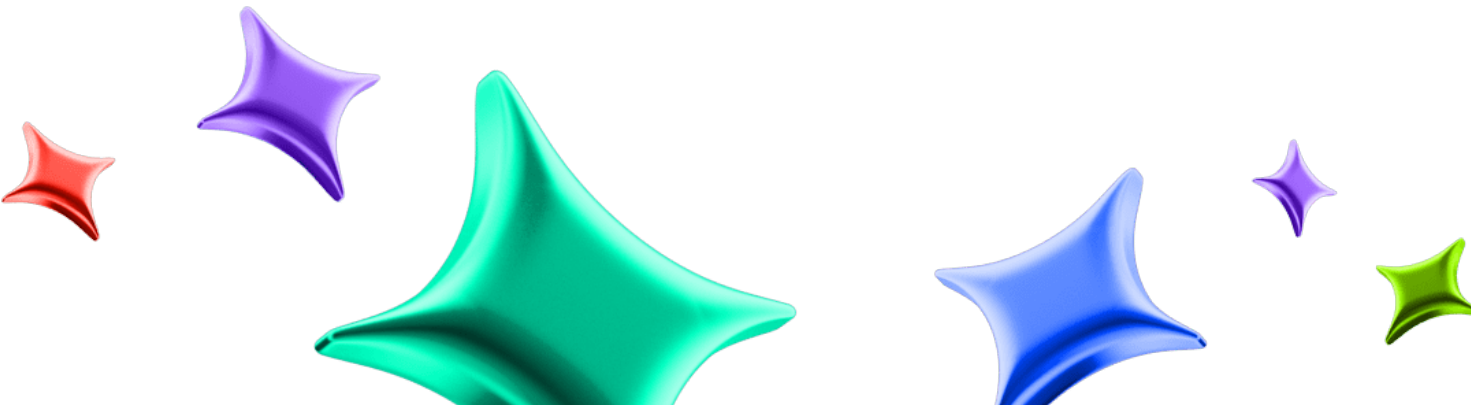


Пространство решений

Для обучения модели вы можете использовать только данные, предоставленные в тренировочной и тестовой выборке. В своем решении вы можете:

- 01** Обучить модель на части датасета.
- 02** Использовать для обучения модели только часть признаков.
- 03** Использовать сгенерированные в ходе предобработки признаки.
- 04** Проводить селекцию и отбор признаков.

Вы имеете право использовать любую модель из любой Open-Source-библиотеки для предсказания. Выбор и настройка модели не ограничены организаторами.



Приложение 1. Метрика WMAE

Метрика WMAE (Weighted Mean Absolute Error) — это модификация стандартной метрики MAE (Mean Absolute Error), которая используется в задачах регрессии и классификации для оценки точности модели. Отличие WMAE от обычной MAE заключается в том, что ошибки для разных примеров взвешиваются с учетом определенных коэффициентов, что позволяет придавать большее значение ошибкам на определенных данных. Для расчета WMAE используется формула:

$$\text{WMAE} = \frac{1}{\sum w_i} \sum_{i=1}^n w_i |y_i - \hat{y}_i|$$

y_i — истинное значение;

\hat{y}_i — предсказанное значение;

w_i — вес для примера i ;

n — общее количество примеров.

Веса w_i позволяют подчеркнуть важность определенных данных в зависимости от задачи. Если все веса равны 1, WMAE совпадает с обычной MAE.

CHANGELLENGE >>

Задание написано и опубликовано
Changellenge >> —
ведущей организацией
по кейсам в России.

www.changellenge.com

Альфа Банк

Задание создано по заказу
АО «Альфа-Банк»

www.alfabank.ru

