# Отток клиентов

Поздравляем! Вы прошли курс в тренажёре. Самое время проверить знания и решить новую задачу машинного обучения. Выполнять работу вы будете самостоятельно.

Когда закончите, отправьте работу на проверку ревьюеру: он пришлёт комментарии в течение суток. После этого нужно доработать проект и пройти повторную проверку.

Скорее всего, вы будете дорабатывать кейс по комментариям ещё несколько раз. Это нормально.

Проект завершён, когда ревьюер одобрит все доработки.

# Описание проекта

Из «Бета-Банка» стали уходить клиенты. Каждый месяц. Немного, но заметно. Банковские маркетологи посчитали: сохранять текущих клиентов дешевле, чем привлекать новых.

Нужно спрогнозировать, уйдёт клиент из банка в ближайшее время или нет. Вам предоставлены исторические данные о поведении клиентов и расторжении договоров с банком.

Постройте модель с предельно большим значением *F1*-меры. Чтобы сдать проект успешно, нужно довести метрику до 0.59. Проверьте *F1*-меру на тестовой выборке самостоятельно.

Дополнительно измеряйте *AUC-ROC*, сравнивайте её значение с *F1*-мерой.

### Инструкция по выполнению проекта

Загрузите и подготовьте данные. Поясните порядок действий.

Исследуйте баланс классов, обучите модель без учёта дисбаланса. Кратко опишите выводы.

Улучшите качество модели, учитывая дисбаланс классов. Обучите разные модели и найдите лучшую. Кратко опишите выводы.

Проведите финальное тестирование.

### Описание данных

Данные находятся в файле /datasets/Churn.csv (англ. «отток клиентов»). [Скачать датасет](https://code.s3.yandex.net/datasets/Churn.csv)

**Признаки**

*RowNumber —* индекс строки в данных

*CustomerId* — уникальный идентификатор клиента

*Surname —* фамилия

*CreditScore —* кредитный рейтинг

*Geography —* страна проживания

*Gender —* пол

*Age —* возраст

*Tenure —* сколько лет человек является клиентом банка

*Balance —* баланс на счёте

*NumOfProducts —* количество продуктов банка, используемых клиентом

*HasCrCard —* наличие кредитной карты

*IsActiveMember —* активность клиента

*EstimatedSalary —* предполагаемая зарплата

**Целевой признак**

*Exited* — факт ухода клиента

# Как будут проверять проект?

Мы подготовили критерии оценки проекта, которыми руководствуются ревьюеры. Прежде чем приступить к решению кейса, внимательно их изучите.

На что обращают внимание ревьюеры, проверяя проект:

Как вы готовите данные к обучению? Все ли типы признаков обрабатываете?

Хорошо ли поясняете этапы предобработки?

Как исследуете баланс классов?

Изучаете ли модель без учёта дисбаланса классов?

Какие выводы об исследовании задачи делаете?

Корректно ли разбиваете данные на выборки?

Как работаете с несбалансированными классами?

Правильно ли проводите обучение, валидацию и финальное тестирование модели?

Насколько высокое значение *F1*-меры получаете?

Изучаете ли значения метрики *AUC-ROC*?

Следите за структурой проекта и поддерживаете аккуратность кода?

Всё, что вам нужно знать, есть в шпаргалках и конспектах прошлых тем.

Успеха!