**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**

1. Загрузите и подготовьте данные. Поясните порядок действий.
2. Исследуйте баланс классов, обучите модель без учёта дисбаланса. Кратко опишите выводы.
3. Улучшите качество модели, учитывая дисбаланс классов. Обучите разные модели и найдите лучшую. Кратко опишите выводы.
4. Проведите финальное тестирование.

**Описание данных**

Данные находятся в файле /datasets/Churn.csv (англ. «отток клиентов»).

**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

# Выводы по итогу проведенного анализа

Рассматривая данные с учетом дисбаланса + переведенные категориальные + стандартизированные данные и без учета дисбаланса тремя моделями можно сделать вывод, что самая лучшая модель - Случайный лес. При лучшем значении max\_depth=12,n\_estimators=24, мы имеем F1 = 0,601 выборке. Хуже всего себя показала модель Логистическая регрессия.

Тестовая модель показала F1 = 0.639, что выше полученных данных на лучшей предыдущей выборке (0.6601). И выше заданной в условии метрики F1 = 0.59.

Протестированная модель выделила долю верного прогноза по оттоку клиентов = 67% из вариантов, признанных положительными = 61%.

Значит банку в целом, и отделу маркетинга, в частном, нужно пересмотреть принципы организации работы, собрать обратную связь о причинах расторжения договоров. Сделать выводы.