**Для решения задачи "прогнозирование оттока пользователей провайдера телекоммуникационных услуг" применялся поэтапный подход:**

* Был произведен обзор и анализ предоставленных данных.
* Определены и реализованы алгоритмы для формирования признаков из предоставленного набора данных для обучающей выборки (feature extraction & feature engineering).
* Ввиду наличия дисбаланса классов в обучающей выборке, произведена минимизация дисбаланса при помощи оверсэмплинга (oversampling) – метода синтетического создания образцов наименьшего класса.
* Была определена, на основе метода  кроссвалидации, лучшая модель машинного обучения (machine learning).
* В результате обучена на всех данных модель машинного обучения на основе градиентного бустинга (CatBoost), с помощью которой сформирован результат по тестовой выборке.

**Технические особенности:**

при реализации решения использовался следующий стек Open Source технологий: **python, numpy, pandas, matplotlib, seaborn, sklearn, catboost**

**Уникальность:**

Решение реализовано с помощью оригинальных алгоритмов извлечения и формирования признаков для обучающей выборки (feature extraction & feature engineering). Обученная на «обогощенных», в результате оверсэмплинга, данных модель демонстрирует увеличение точности по метрике Recall более чем на 15% по сравнению с точностью модели, обученной на данных без добавления синтетически созданных элементов меньшего класса.

**Масштабируемость решения:**

Возможно многократно увеличить данные для обучения модели, так как на датасете менее 7 тысяч записей всё решение, включая извлечение признаков, формирование обучающей выборки, её оверсэмплинг и непосредственно обучение модели занимает 1 минуту.