

ПРОГНОЗИРОВАНИЕ ОТТОКА КЛИЕНТОВ

Щирова Анна

ЗАДАЧИ И ЦЕЛИ ПРОЕКТА

- **Задача проекта** - Прогнозирование вероятности того, что клиент перестанет пользоваться сервисом в течение некоторого заданного заранее промежутка времени.
- **Цель проекта** - научиться находить пользователей, склонных к оттоку. Если научиться находить таких пользователей с достаточной точностью заблаговременно, то можно эффективно управлять оттоком.

Задача прогнозирования оттока является одной из важнейших подзадач в области работы с аудиторией и актуальна не только для телекоммуникационных компаний, но и для большинства организаций, оказывающих услуги в сегменте B2C. Такие задачи часто возникают на практике у телекоммуникационных операторов, провайдеров кабельного телевидения, страховых компаний, банков, крупных и средних интернет-порталов и др.

МЕТОДИКА ИЗМЕРЕНИЯ КАЧЕСТВА И КРИТЕРИЙ УСПЕХА

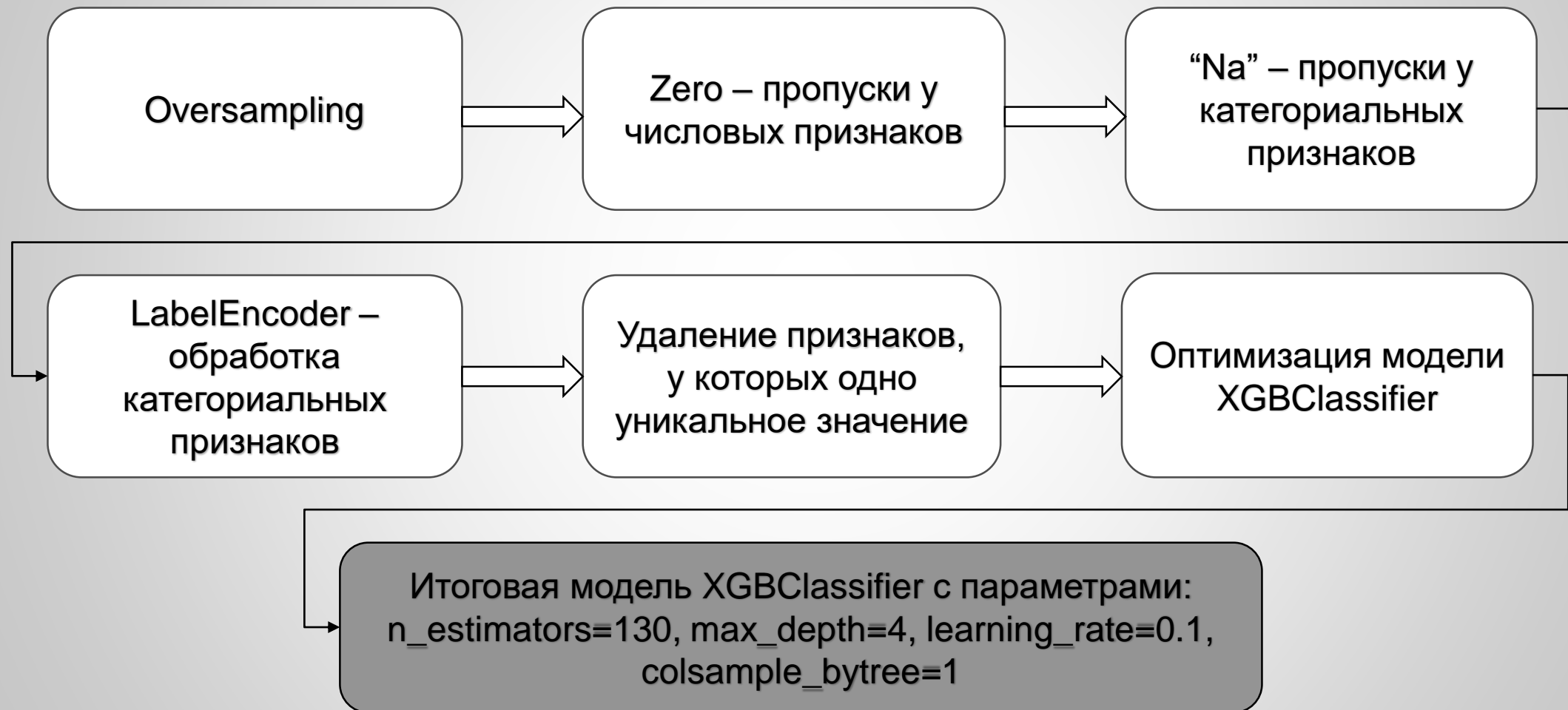
- Для оценки качества использовались метрики ROC-AUC и RECALL.
- Оценка производилась по CV (nfold = 5) и Hold out.
- Критерием успешности считаем значения метрик качества выше baseline.
- Производилась оценка экономического эффекта от внедрения модели.

ТЕХНИЧЕСКОЕ ОПИСАНИЕ РЕШЕНИЯ

Проведенные эксперименты:

- Балансировка: oversampling, undersampling, разные веса
- Обработка пропусков: mean, zero, median, аномальные значения
- Обработка категориальных признаков: Label encoding, One-hot encoding, замена каждой категории числом входящих в неё объектов
- Отбор признаков: Feature importances, SelectKBest, удаление признаков с большим числом пропусков, удаление категориальных признаков с большим числом различных категорий
- Модели: LogisticRegression, RandomForestClassifier, GradientBoostingClassifier, XGBClassifier

В результате описанных выше экспериментов итоговая модель с наилучшим качеством была построена следующим образом:



ВЫВОДЫ О КАЧЕСТВЕ МОДЕЛИ

Качество итоговой модели:

- Основная метрика: AUC-ROC
- Кроссвалидация (CV) на 5 фолда
- Оценка на CV: AUC-ROC 0.874113
- Итоговая оценка AUC-ROC на отложенной выборке: 0.79909

ОЦЕНКА ЭКОНОМИЧЕСКОГО ЭФФЕКТА ОТ ВНЕДРЕНИЯ МОДЕЛИ

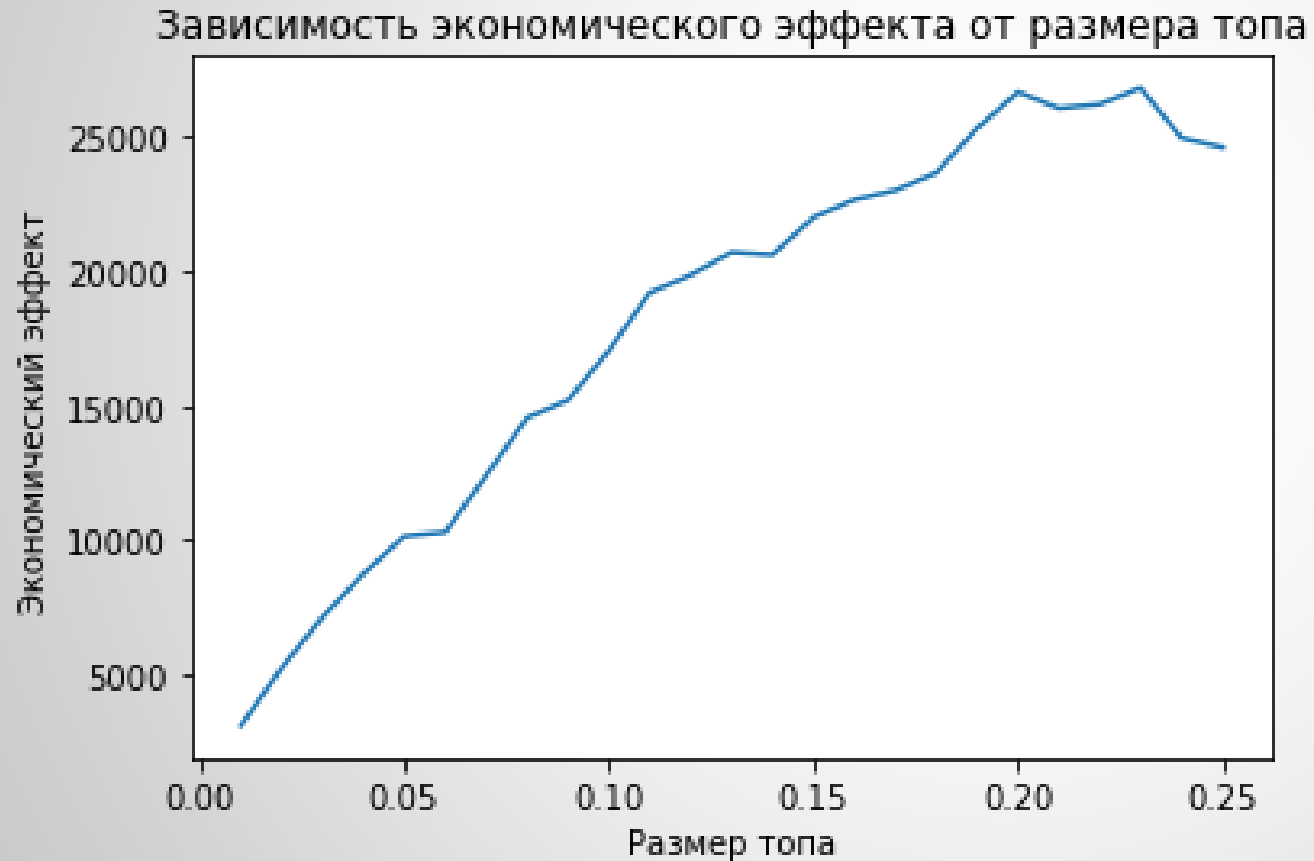
Введем следующие переменные:

- IOC (Income one customer) - сколько денег в среднем приносит один пользователь в месяц
- PR (Priese for hold) - сколько денег в среднем будем вкладывать в удержание одного пользователя
- PA (Probability accept) - с какой вероятностью пользователь примет наше предложение
- TP (Top customers) - сколько пол-й согласно ранжированию по модели будет участвовать в кампании по удержанию
- TC (True-true churn) - число польз-й, к-ых модель отнесла к классу «отток» и кот-ые действительно собирались уйти
- FC (False churn) - число польз-й, к-ых модель отнесла к классу «отток», но кот-ые по факту не собирались уходить
- FnoC (False no-churn) – число польз-й, к-ых модель отнесла к классу «не отток», но кот-ые по факту хотят уйти

Размер экономического эффекта будем вычислять по формуле:

$$Ef \equiv (IOC*PA-PR)*TC - PR*FC - IOC*FnoC$$

ЗАВИСИМОСТЬ ЭКОНОМИЧЕСКОГО ЭФФЕКТА ОТ РАЗМЕРА ТОПА



Максимальный размер экономического эффекта достигается при применении кампании по удержанию к 23% топу согласно ранжированию модели.

ДАЛЬНЕЙШАЯ РАБОТА

- Улучшение качества прогноза модели.

Улучшение качества прогноза (recall) на 1% дает оценку увеличения прибыли кампании примерно на 10%

- Проведение A/B тестирования.

Можно провести следующим образом: случайным образом сформировать 2 группы клиентов – контрольную и тестовую. В тестовой проводить мероприятия по удержанию согласно прогнозу модели. Через 3 месяца сравнить результаты с помощью критерия Манна - Уитни

ПОДВЕДЕНИЕ ИТОГОВ

В рамках финального проекта специализации «Машинное обучение и анализ данных» была построена модель прогнозирования оттока клиентов и проделаны следующие шаги:.

- Проведен первичный анализ данных.
- Была проведена предобработка данных (работа с пропусками, с категориальными признаками, балансировка классов, отбор признаков).
- Были рассмотрены модели из различных групп (линейные, деревья).
- Произведен подбор параметров модели.

Выводы:

- Оценка экономического эффекта показала, что внедрение модели в деятельность компании имеет смысл и экономическую выгоду.
- В дальнейшем необходимо развитие и улучшение модели, поскольку это может привести к еще большей доходности от использования модели.