



# Прогнозирование оттока клиентов

Губжоков Алим Лиуанович

04 января 2019 г.

# Содержание

- ▶ Цели и задачи проекта
- ▶ Методика измерения качества и критерий успеха
- ▶ Техническое описание решения
- ▶ Выводы о качестве модели, ожидаемый экономический эффект от использования модели
- ▶ Подведение итогов

# Цели и задачи проекта

- ▶ Отток пользователей - одна из наиболее актуальных задач в областях, где распространение услуги составляет порядка 100%. Ярким примером такой области является сфера телекома. В связи с чем, важным является удержание клиентов, а не привлечение новых.
- ▶ Решаемая задача - бинарной классификации: отток и не отток.
- ▶ Разработанная модель прогнозирования позволит находить клиентов, склонных к оттоку по их поведению и для проведения последующих мероприятий (обзвон клиентов и предложение более выгодного тарифного плана или скидка на текущий тарифный план), связанных с удержанием клиентов.

# Методика измерения качества и критерий успеха

- ▶ Из-за несбалансированности классов имеет смысл использовать метрику ROC-AUC.
- ▶ Метрики такие как PR-AUC, accuracy, precision, recall смысла использовать нет, так как эти метрики очень чувствительны к несбалансированности классов.
- ▶ Процесс тестирования модели лучше проводить на новых данных, возможно АВ-тестирование.
- ▶ Тестировать модель лучше всего рассматривая экономический эффект, так как экономический эффект от внедрения модели будет показывать реальный результат работы, так же это позволит получить детальную информацию о цене ошибок первого и второго рода.
- ▶ Критерием успешности можно считать увеличение площади ROC-кривой на 15-20% и увеличение экономического эффекта после АВ-тестирования.

# Техническое описание решения

## ЭКСПЕРИМЕНТАЛЬНО ИСПОЛЬЗОВАННЫЕ СТРАТЕГИИ

1

БАЛАНСИРОВКА  
ВЫБОРКИ

Undersampling

Oversampling

2

ОБРАБОТКА  
КАТЕГОРИАЛЬНЫХ  
ПРИЗНАКОВ

Label Encoding

One-hot  
Encoding

Frequency  
Encoding

3

ЗАПОЛНЕНИЕ  
ПРОПУСКОВ

Mean

Zeros

Most

Frequent

4

РЕГУЛЯРИЗАЦИЯ

Lasso

Ridge

6

ПОДБОР  
ГИПЕРПАРАМЕТРОВ

GridSearchCV

5

КЛАССИФИКАТОРЫ

Ridge  
Classifier

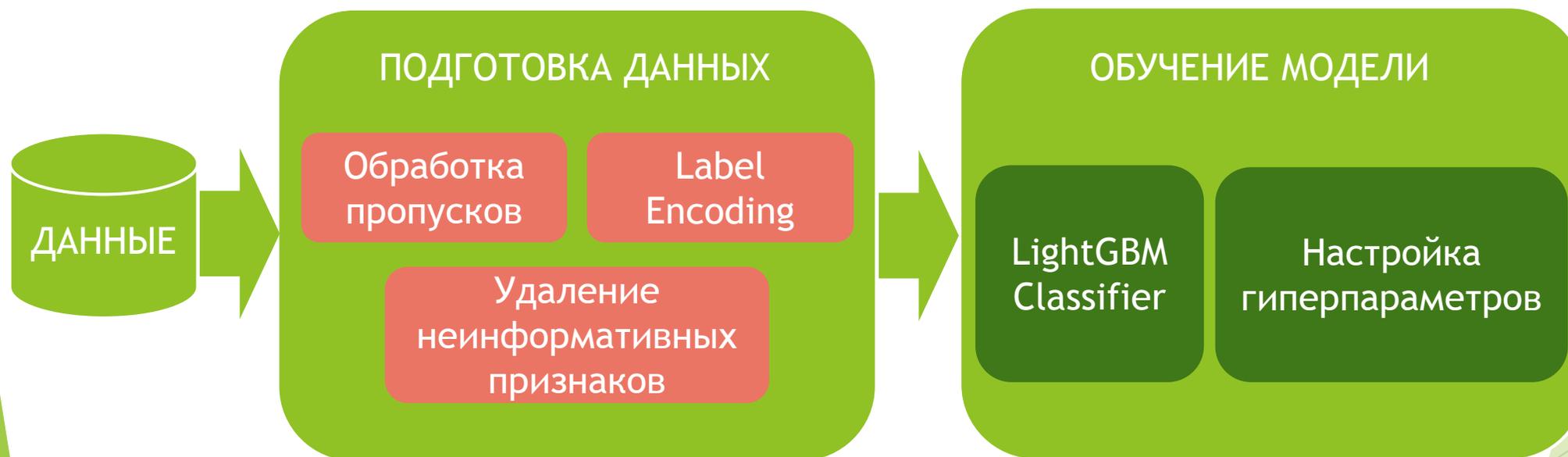
Logistic  
Regression

RandomForest  
Classifier

GradientBoosting  
Classifier

# Техническое описание решения

## СХЕМА ЭТАПОВ ОБРАБОТКИ ДАННЫХ



# Выводы о качестве модели

## Ожидаемый экономический эффект от использования модели

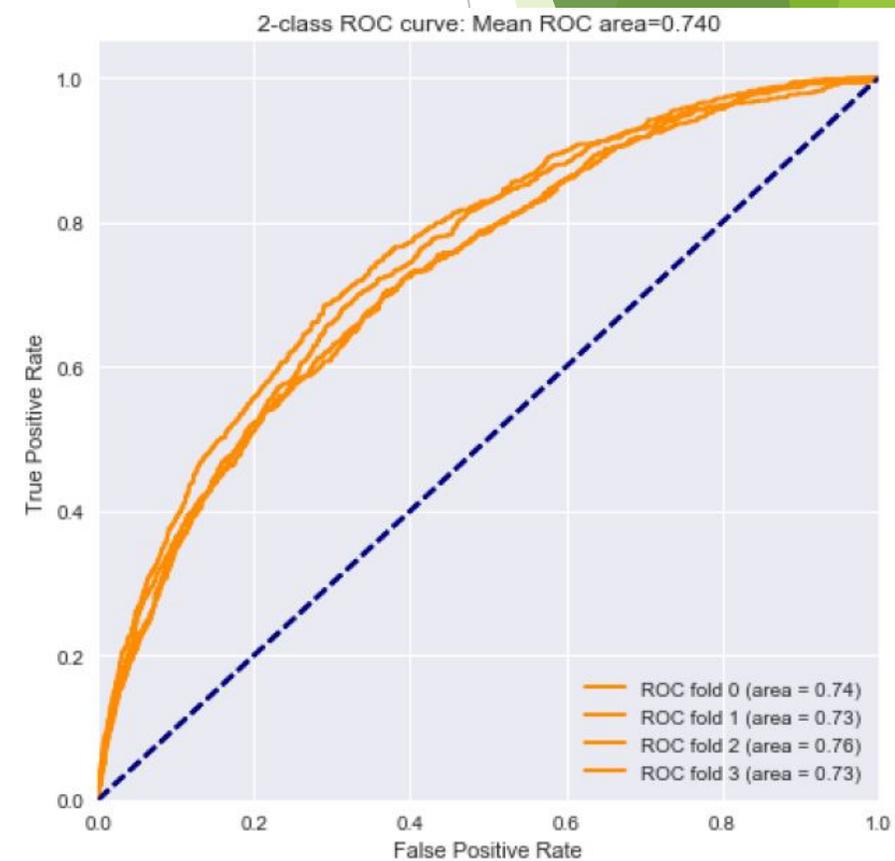
- ▶ **МЕТРИКА:**
- ▶ **ИТОГОВАЯ ОЦЕНКА:**
- ▶ **КРОССВАЛИДАЦИЯ НА ОБУЧЕНИИ:**
- ▶ **AUC SCORE (НА ОТЛОЖЕННОЙ ВЫБОРКЕ):**
- ▶ **ТОП-6 ПРИЗНАКОВ С НАИБОЛЬШИМ ВКЛАДОМ В МОДЕЛЬ:**

**AUC-ROC**

на отложенной  
выборке  
4 фолда

**0.738164**

**Var125, Var112,  
Var216, Var188,  
Var73, Var209**



# Подведение итогов

- ▶ РАЗМЕР ОПТИМАЛЬНОГО ТОПА ПРИ КОТОРОМ ЭКОНОМИЧЕСКИЙ ЭФФЕКТ МАКСИМАЛЬНЫЙ **10-12%**
- ▶ ОБУЧЕННАЯ МОДЕЛЬ ПОЛОЖИТЕЛЬНЫЙ ИМЕЕТ ЭКОНОМИЧЕСКИЙ ЭФФЕКТ - ПРИБЫЛЬ ~ **4000Р** ПО СРАВНЕНИЮ С ГРУППОЙ КЛИЕНТОВ, НАД КОТОРЫМИ КАМПАНИЯ НЕ ПРОВОДИЛАСЬ.
- ▶ МИНИМАЛЬНУЮ ПРИБЫЛЬ ПОКАЗЫВАЕТ ГРУППА, НАД КОТОРОЙ ПРОВОДИЛАСЬ ПОЛНАЯ КАМПАНИЯ ПО УДЕРЖАНИЮ.