

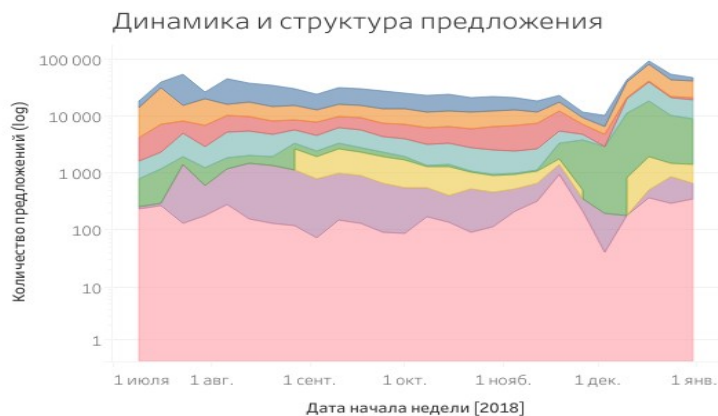
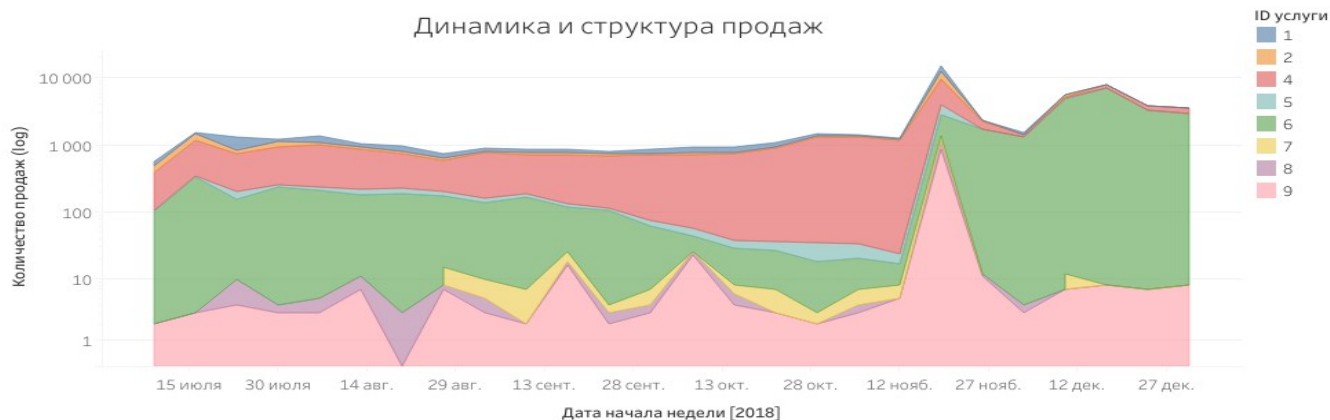
Прогнозирование вероятности подключения услуги абонентом сети «Мегафон»

Разработал и подготовил
Петр Мирзоян

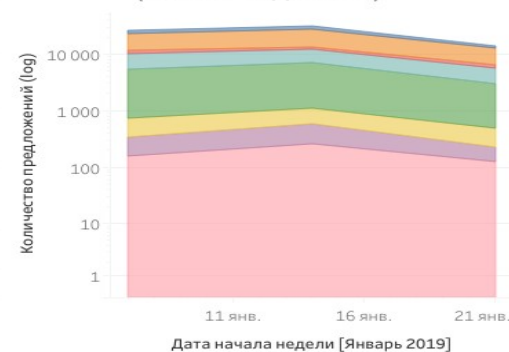
Описание задачи и входные данные

- Задача
 - определить вероятность подключения услуги для пары *пользователь-услуга*
- Данные
 - информация об отклике абонента на предложение
 - анонимизированные признаки профиля абонента, включающие дату агрегации

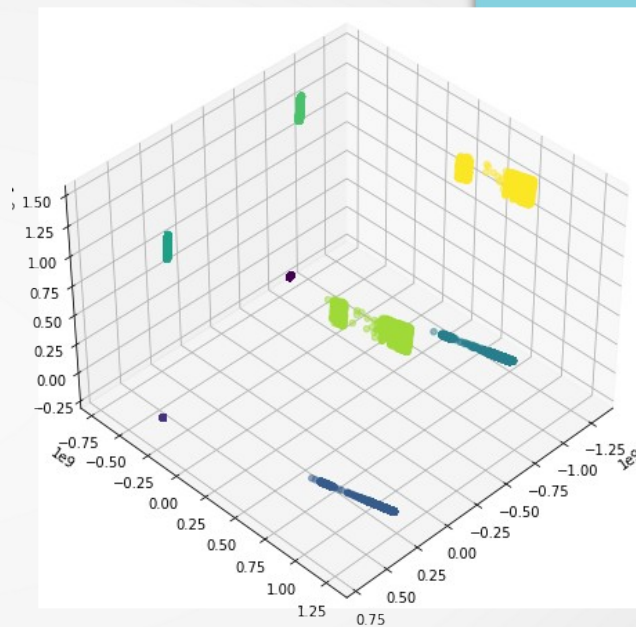
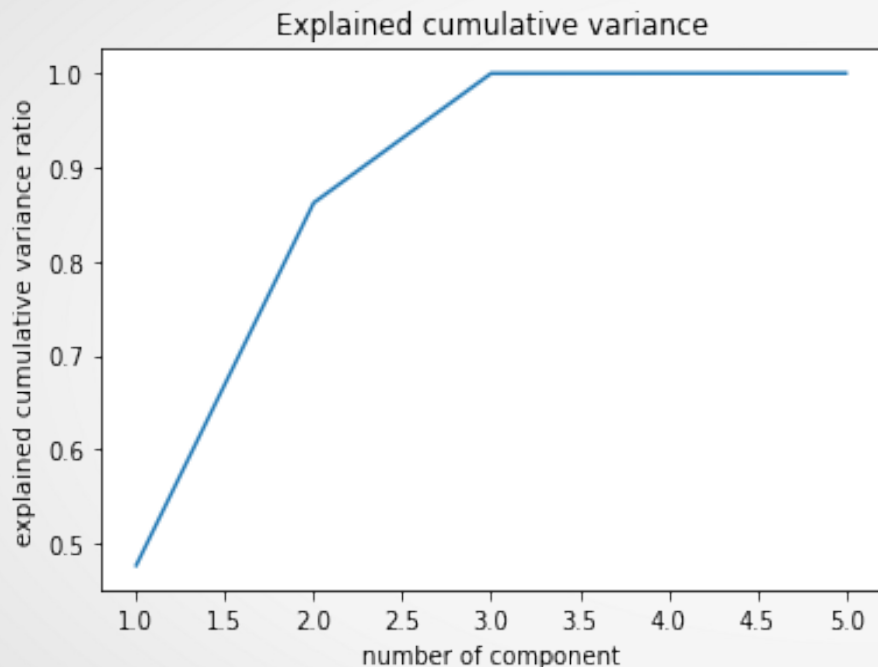
Основа модели: структура спроса и предложения



Динамика и структура предложения
(тестовые данные)



РСА сжатие признаков пользователей



- Количество признаков можно сократить до 3 практически без потери информации
- Сжатые признаки образуют явно различимые 8 кластеров

Сравнение моделей

Способ сопоставления	SGD	GradientBoosting	RandomForest	LGBM	XGBoost
NEAREST	0.449580	0.551522	0.720384	0.737395	0.619565
BACKWARD	0.425094	0.586954	0.739461	0.753410	0.644542
BACKWARD + NEAREST	0.437933	0.550974	0.720716	0.737490	0.619077

RandomForest

- показывает хорошую метрику на старте
- не требует дополнительных компонентов в образе airflow

Ключевые особенности

- Автоматизация процессов (airflow)
 - обучения и валидации
 - подбора гиперпараметров
- Не требует распределенных вычислений для прогноза
- Простота настройки

Конфигурация модели

- Параметры модели (`parameters.conf`)
 - список игнорируемых признаков пользователей
 - дата «отсечки» обучающих данных
 - кол-во разбиений для кросс-валидации
- Параметры обучения (`fit_params.json`)
 - стандартные параметры `RandomForest`

Составление индивидуальных предложений

- Получить прогноз модели для каждой пары *пользователь-услуга*
- Отбросить услуги с прогнозной вероятностью ниже условленного порога
- Отсортировать услуги в порядке убывания прогнозной вероятности

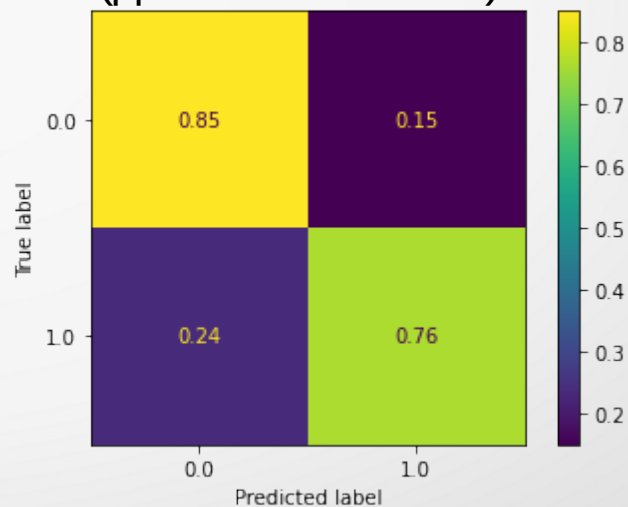
Результаты

Средняя метрика (f1_score) на 5 разбиениях = 0.735381

Правильно классифицированы

- 76% подключивших услугу
- 85% не подключивших услугу
- 84% всех объектов

Матрица ошибок
(доли по классам)

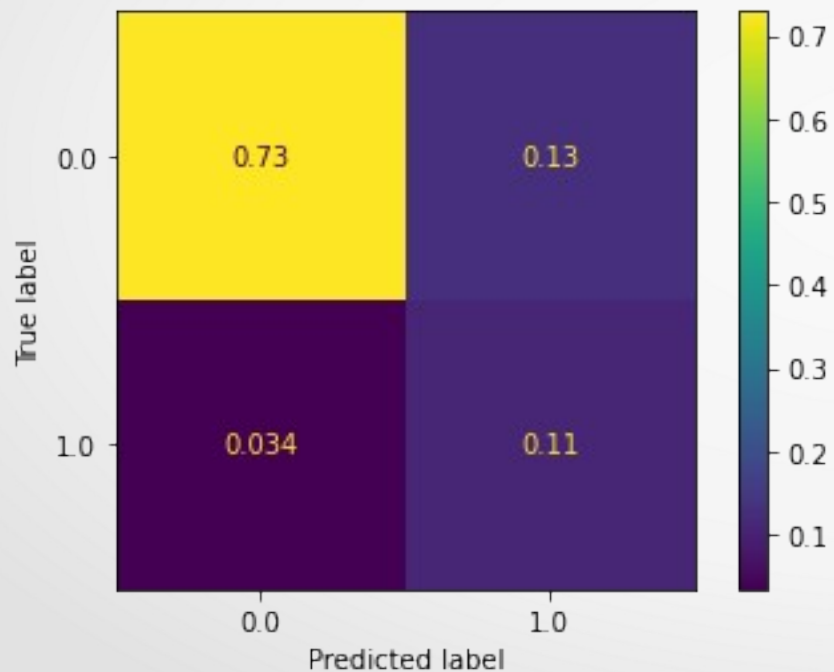




Спасибо за внимание

Дополнительный материал

Матрица ошибок
(доли от общего)



Матрица ошибок
(доли по классам)

