



# НАСКАТНОН BEST URFU 2019



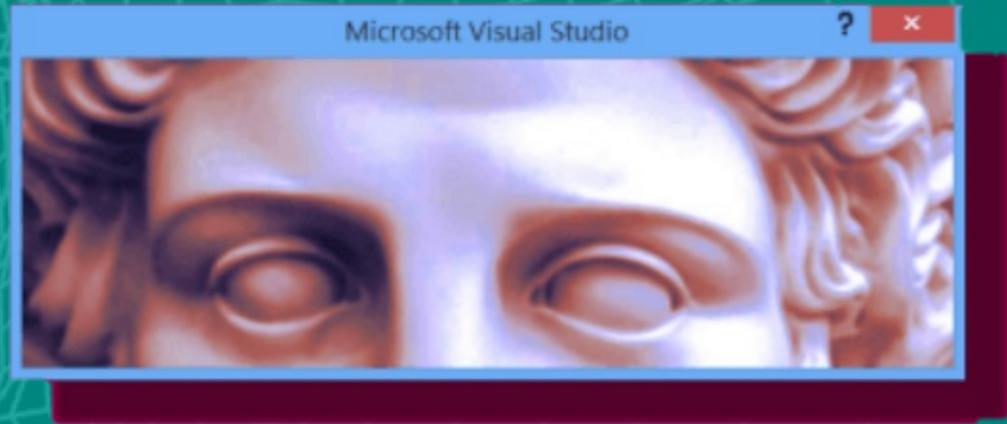
Отборочное  
задание  
**Data Science**



# РАЗРАБОТКА И ВНЕДРЕНИЕ МОДЕЛИ ВЫЯВЛЕНИЯ АБОНЕНТОВ, СКЛОННЫХ К РАЗВИТИЮ (CROSS SELL)

1. Использованные понятия и сокращения .....	3
2. Цель моделирования .....	3
3. Задачи моделирования .....	4
3.1. Определение Cross Sell и подхода к его моделированию .....	4
3.2. Формирование генеральной совокупности.....	4
3.3. Идентификация статуса абонентов .....	4
3.4. Формирование выборки.....	4
3.5. Аудит и очистка данных.....	4
3.6. Подготовка предикторов .....	4
3.7. Отбор предикторов и настройка модели.	4
4. Результаты моделирования .....	5
5. Передача результатов моделирования .....	5
6. Приемка результатов моделирования .....	5





# ОСНОВНЫЙ ПОНЯТИЯ И ЦЕЛИ МОДЕЛИРОВАНИЯ

Термин	Определение
Cross Sell	Приобретение абонентом новой для него телекоммуникационной Услуги
Генеральная совокупность	Совокупность абонентов, удовлетворяющая заданным критериям и используемая для построения, применения, оценки качества модели и формирования статистических выводов
Сбалансированная выборка	Выборка, в которой в равной степени представлены все категории статуса абонента, то есть для двух категорий 50% абонентов находятся в категории 0 и 50% абонентов в категории 1

**Целью моделирования** является получение инструмента диагностики абонентов, склонных к подключению услуги.

Абоненты с высокой вероятностью подключения Услуги в предстоящем периоде в дальнейшем должны включаться в целевые Cross Sell - кампании по предоставлению данной услуги.

# ЗАДАЧИ МОДЕЛИРОВАНИЯ

## 3.1. Определение Cross Sell и подхода к его моделированию

Cross Sell должно считаться приобретение абонентом новой для него телекоммуникационной Услуги Заказчика.

Моделированием Cross Sell должен считаться процесс прогнозирования вероятности подключения новой Услуги абонентом из заданной генеральной совокупности на интервале прогноза, равном по продолжительности интервалу определения статуса.

Моделирование должно включать создание «Модели Cross Sell» путем исследования взаимосвязей между характеристиками и текущим поведением абонента в отношении потребления Услуги и статусом наличия Услуги у абонента на заданном будущем временном интервале.

Исследования, анализ и прогнозирование должны осуществляться на исторических данных, характеризующих абонентов и их поведение.

Исследования поведения абонента должны осуществляться на 2-х интервалах: ретроспективном - интервале анализа и перспективном - интервале прогноза.

Достижение цели моделирования должно осуществляться последовательной реализацией двух целевых задач: настройка «Модели Cross Sell» и применение «Модели Cross Sell» (скоринг).



Рисунок 1. Настройка «Модели Cross Sell»

- Основными выполняемыми задачами должны быть:
- Определение генеральной совокупности абонентов
  - Идентификация статуса абонента
  - Настройка «Модели Cross Sell» на основе исторических данных
  - Тестирование настроек модели и интерпретация модели
  - Скоринг «Модели Cross Sell» на основе текущих данных

Результатами моделирования Cross Sell должны являться итоги процесса прогнозирования подключения Услуги (скоринги), применимые при планирования целевых Cross Sell - кампаний.

# ЗАДАЧИ МОДЕЛИРОВАНИЯ

## 3.2. Формирование генеральной совокупности

Задача должна быть реализована в рамках требований к входным данным, ограничениям и условиям, формату и объему результатов, определенных ниже.

Назначением задачи должно быть определение критериев включения абонента в исследуемое множество.

Для состоятельности целевых результатов, их сквозной согласованности и целостности, настройка, обучение и применение «Модели Cross Sell» должны выполняться на основании единых критериев формирования генеральной совокупности.

Критерии включения абонента в исследуемое множество должны соответствовать бизнес-целям скоринга. Критерии должны быть выбраны в процессе настройки «Модели Cross Sell».

## 3.3. Идентификация статуса абонентов

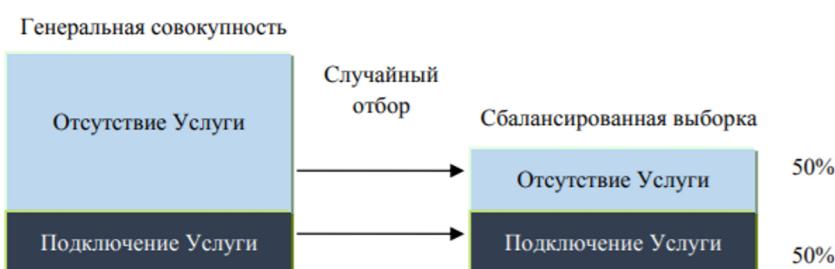
Статус 1 (подключение Услуги) должен определяться как событие, когда на временном интервале определения статуса абонент подключил Услугу. Статус 0 (отсутствие подключения Услуги) должен определяться как событие, когда на временном интервале определения статуса абонент не подключил Услугу.

## 3.4. Формирование выборки

Задача должна быть реализована в рамках требований к входным данным, ограничениям и условиям, формату и объему результатов, определенных ниже.

Назначением задачи должно быть формирование перед настройкой «Модели Cross Sell» сбалансированной выборки для каждой исследуемой генеральной совокупности абонентов.

Сбалансированная выборка должна обеспечить сопоставимость оценок точности моделей при первой настройке «Модели Cross Sell» и последующих ее обновлений, в том числе выравнивание влияния предикторов абонентов, подключивших и неподключивших Услугу для корректного различия паттернов трафика этих групп.



Конечным результатом задачи должна быть сбалансированная выборка абонентов для генеральной совокупности абонентов, исследуемых далее.

Рисунок 4. Схема формирования выборки для «Модели Cross Sell»

# **ЗАДАЧИ МОДЕЛИРОВАНИЯ**

## **3.5. Аудит и очистка данных**

Назначением задачи должно быть формирование перед применением «Модели Cross Sell» очищенной сбалансированной выборки из исследуемой генеральной совокупности абонентов.

При аудите данных должны быть выявлены возможные проблемы с данными, влияющие на качество результатов применения «Модели Cross Sell».

Задача должна выполняться параллельно в ходе решений всех задач настройки и применения «Модели Cross Sell».

Конечным результатом задачи должна быть сбалансированная выборка абонентов для исследуемой генеральной совокупности абонентов, качество данных которых улучшено.

## **3.6. Подготовка предикторов**

Назначением задачи должно быть формирование перед применением «Модели Cross Sell» ее потенциальных предикторов.

В ходе решения задачи должны быть выполнены расчеты значений производных показателей «Модели Cross Sell».

Потенциальные предикторы должны быть рассчитаны при настройке «Модели Cross Sell».

Конечным результатом задачи должны быть множество показателей потенциальных предикторов «Модели Cross Sell» для исследуемой генеральной совокупности абонентов.



# ЗАДАЧИ МОДЕЛИРОВАНИЯ

## 3.7. Отбор предикторов и настройка модели

Назначением задачи должно быть формирование перед применением «Модели Cross Sell» ее целевых предикторов на основании предварительного множества показателей потенциальных предикторов.

Отбор предикторов должен быть выполнен итерационно в ходе настройки «Модели Cross Sell» с использованием комбинации средств автоматического отбора и поэтапного анализа промежуточных результатов применения «Модели Cross Sell».

При отборе должен быть выполнен анализ точности результатов применения «Модели Cross Sell» на основе случайного разделения выборки на обучающую и тестовую.

Должен быть рассмотрен и выполнен выбор наилучшего алгоритма, в том числе логистическая регрессия, нейронная сеть и т. д.

Конечным результатом задачи должны быть множество показателей целевых предикторов «Модели Cross Sell» и настроенная «Модель Cross Sell» для исследуемой генеральной совокупности абонентов.



# РЕЗУЛЬТАТЫ

---

## 4. Результаты моделирования

Результатами моделирования должны быть:

1. «Модель Cross Sell», настроенная для исследуемой генеральной совокупности абонентов.
2. Документация по ее использованию.

## 5. Передача результатов моделирования

В Отчет должно быть включено описание бизнес-логики предикторов (интерпретацию) в разработанной «Модели Cross Sell».

## 6. Приемка результатов моделирования

Приемка качества «Модели Cross Sell» будет производиться путем сравнения точности модели на независимой тестовой выборке с точностью, полученной при настройке модели.

