

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ #2. РАЗБОР КЕЙСОВ



Максим Чикуров

Data Scientist и руководитель команды
аналитики

Работал в компаниях Citibank, BNP Paribas,
Barclays Bank, Teradata



maxim.chikurov@gmail.com

**О ЧЕМ ПОГОВОРИМ
И ЧТО СДЕЛАЕМ**

О ЧЕМ ПОГОВОРИМ И ЧТО СДЕЛАЕМ

План занятия

- Практика
- Машинные методы обработки данных



ПРАКТИКА

Датасет №2: трафик в интернет-магазине

Столбец №1 – clientID (идентификатор клиента).

Далее: данные по транзакциям, количеству сеансов, виды трафика по устройствам, по каналам, и данные о промежуточных конверсиях (значения 1 и 0). Строки с одним clientID могут повторяться, если любой другой параметр отличен от других строк с этим clientID.

goo.gl/tUkN6j

Датасет №2: трафик в интернет-магазине

Вопросы:

- как соотносятся промежуточные конверсии с транзакциями
- как влияют устройства на промежуточные конверсии и транзакции
- как влияют источники трафика на промежуточные конверсии и на транзакции
- есть ли какая-то связь между количеством сеансов, длительностью сеансов и промежуточными конверсиями и на транзакциями
- можно ли спрогнозировать конверсии и транзакции, исходя из предыдущих данных
- какие характеристики присущи клиентам, которые вообще никогда не закажут

Датасет №3: конверсия из гостевых посещений в клиентские карты

Датасет содержит данные о:

- времени и месте открытия новых клиентских карт тайм-кафе
- времени начала и окончания гостевого посещения тайм-кафе

goo.gl/KNvPSk

Задача: поискать инсайты в данных и сравнить результативность работы кафе

Датасет №3: конверсия из гостевых посещений в клиентские карты

Вопросы:

- как соотносятся конверсии гостевых визитов в карты в ночную, утреннюю и вечернюю смену
- как влияют уменьшение количества кафе на процент конверсии
- какова конверсия из гостевых посещений в клубные карты

* В антикафе клубная система - у гостей есть клубная карта, если они приходят больше 1 раза, либо гостевая карта, если гость приходит в первый раз. В антикафе платят за время, проведенное в клубе. В июне 2018 из 3 клубов закрылся 1, из выгрузки он в этот момент пропадает.

МАШИННЫЕ МЕТОДЫ ОБРАБОТКИ ДАННЫХ

Machine Learning

Машинное обучение (англ. *Machine Learning*) — обширный подраздел искусственного интеллекта; математическая дисциплина, использующая разделы математической статистики, численных методов оптимизации, теории вероятностей, дискретного анализа, и извлекающая знания из данных.

Два типа обучения

- Обучение по прецедентам, или индуктивное обучение, основано на выявлении закономерностей в эмпирических данных
- Дедуктивное обучение предполагает формализацию знаний экспертов и их перенос в компьютер в виде базы знаний

Дедуктивное обучение принято относить к области экспертных систем, поэтому термины *машинное обучение* и *обучение по прецедентам* можно считать **синонимами**

Основные задачи и методы ML

- Регрессия (regression)
- Классификация (classification)
- Кластеризация (clusterization)
- Обучение с подкреплением (reinforcement learning)
- Ассоциативные правила (association rule learning)

СПАСИБО ЗА ВНИМАНИЕ