#mlcourse_open

Открытый курс OpenDataScience и Mail.ru Group
по машинному обучению

Юрий Кашницкий
Программист-исследователь Mail.ru Group

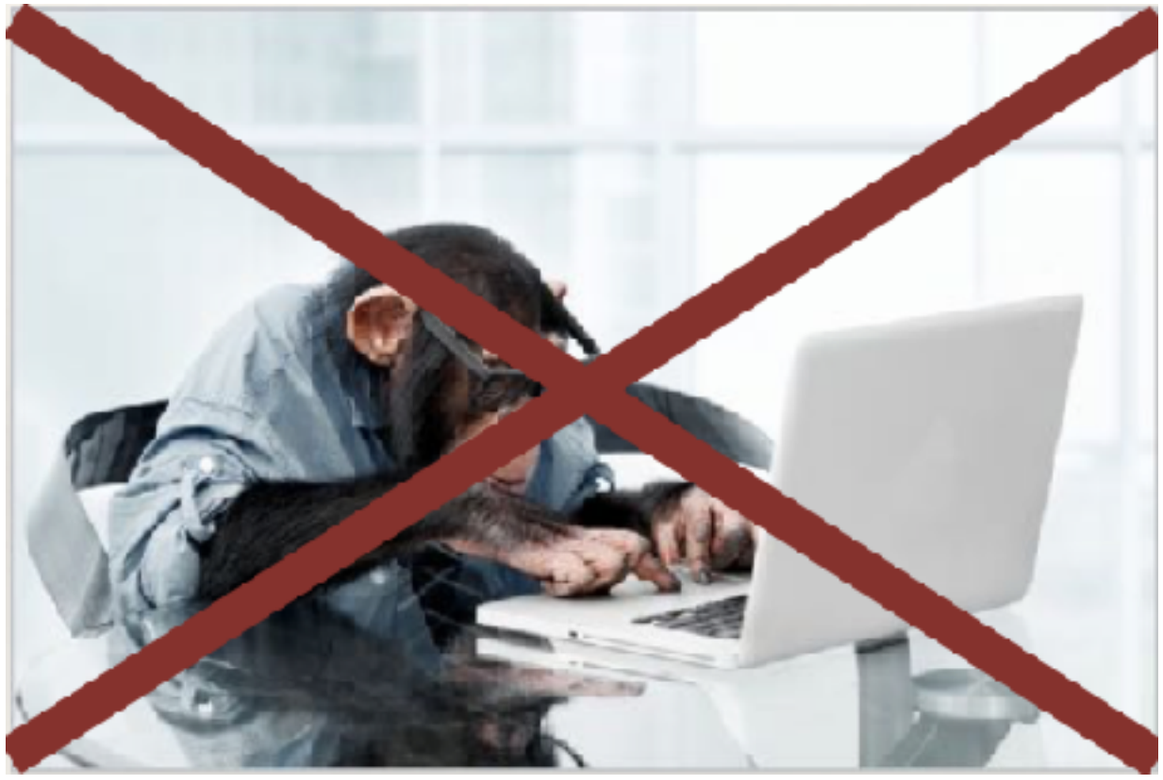
Что нас ждет

Обзор курса

- 10 лекций
- Основные алгоритмы и их использование
- Домашние задания и практики
- Соревнования
- Индивидуальные проекты
- Куча общения
- И не только

Особенности курса

- Обилие практики - задания на каждом занятии и после него
- Теоретическое понимание алгоритмов
- Знакомство с соревнованиями по анализу данных
- Собственный проект
- Community!



Логистика

- Все общение – в Slack ODS. http://ods.ai/
- За домашние задания max 10 баллов
- За проекты, соревнования и тьюториалы – max 40 баллов
- Текущий рейтинг тут https://goo.gl/cd8hUc
- Все материалы курса - на GitHub (mlcourse_open)
- Топ-100 участников будут поощрены

Инструменты

- Язык Python
- Jupyter notebooks
- GitHub
- Docker (опц.)
- Сторонние либы типа Vowpal Wabbit

Занятие 1

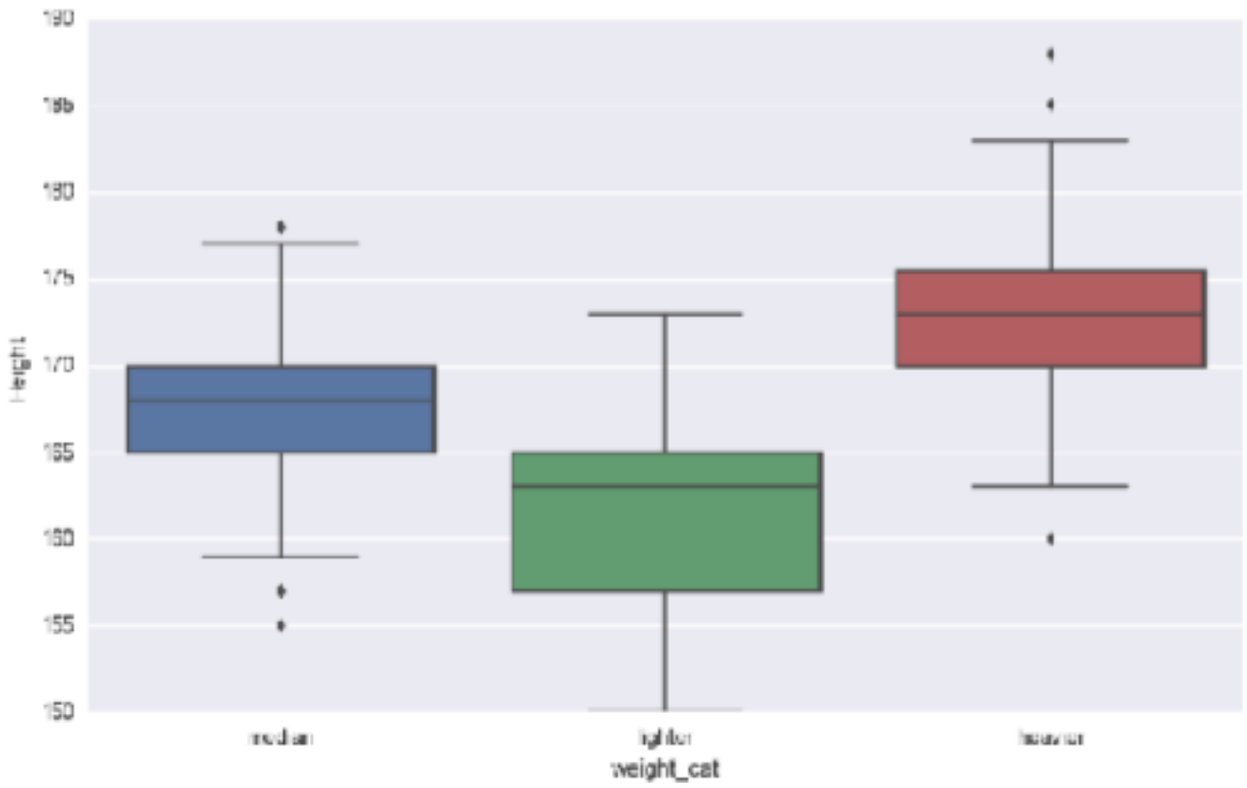
- Анализ данных с Pandas
- Практика на знакомство с данными

```
df.head(4)
```

	wage	exper	union	goodhlth	black	female	married
0	5.73	30	0	1	0	1	1
1	4.28	28	0	1	0	1	1
2	7.96	35	0	1	0	1	0
3	11.57	38	0	1	0	0	1

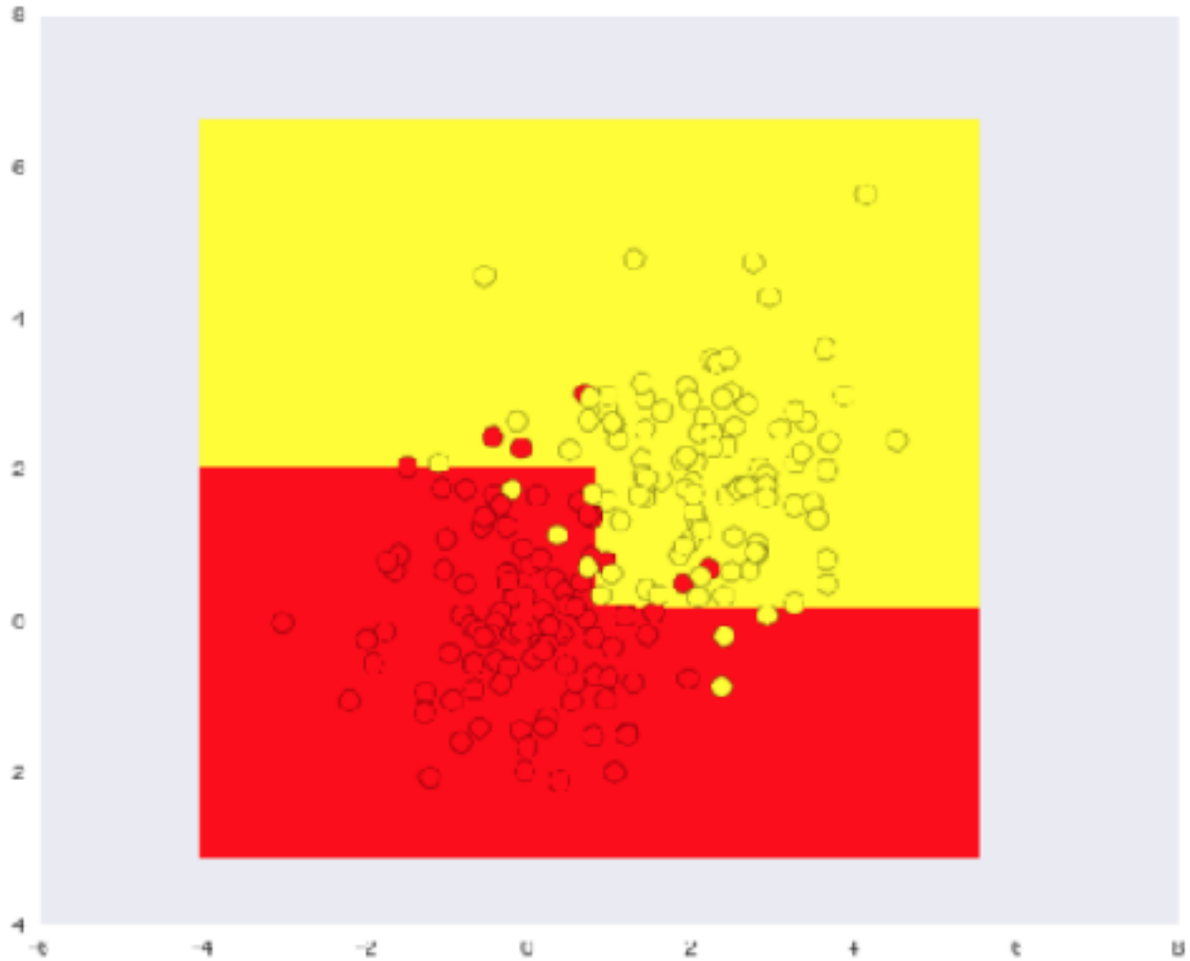
Занятие 2

- Визуальный анализ данных с Pandas и Seaborn
- Практика на «рисование»



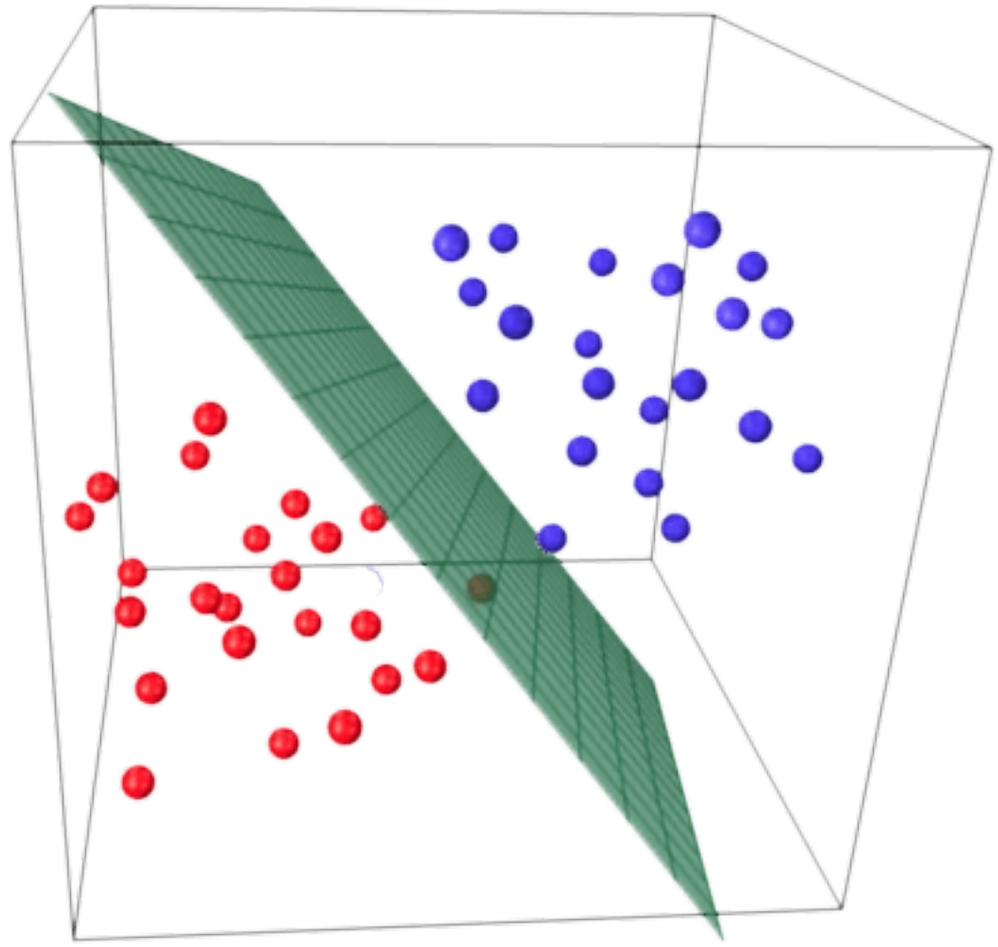
Занятие 3

- Основы машинного обучения
- Деревья решений
- Метод ближайших соседей
- Практика на знакомство с библиотекой Scikit-learn



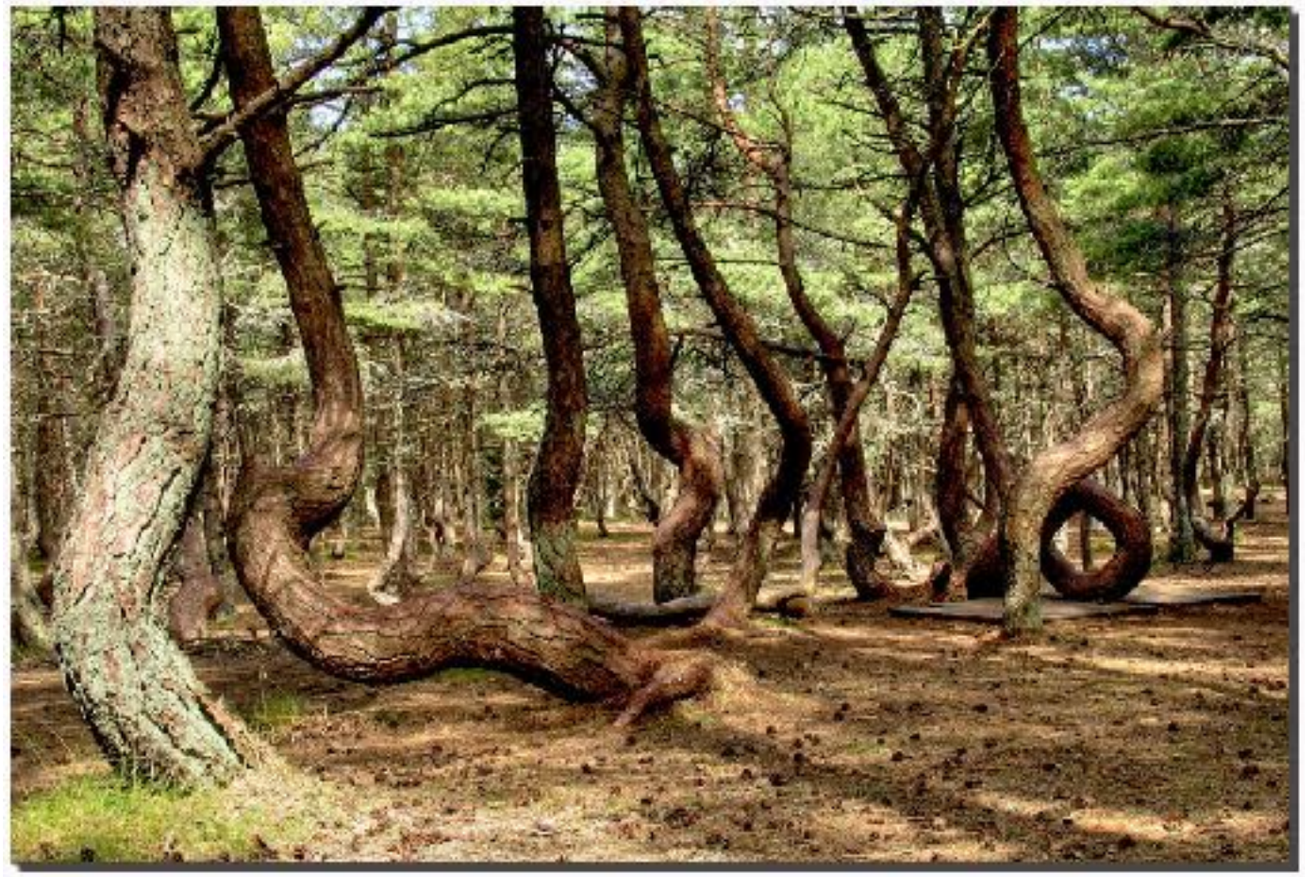
Занятие 4

- Линейные модели классификации
- Регуляризация
- Кросс-валидация
- Практика на логистическую регрессию



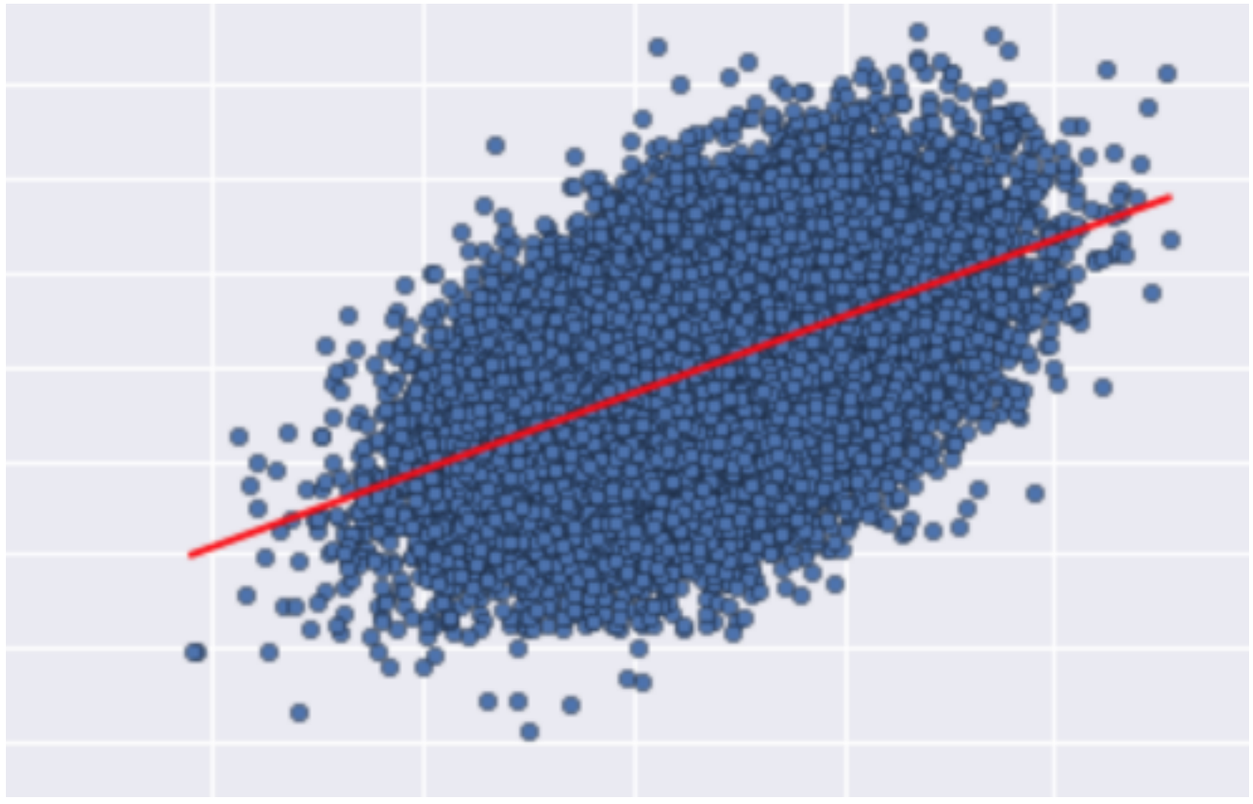
Занятие 5

- Композиции алгоритмов, случайный лес
- Практика на применение случайного леса и оценке важности признаков



Занятие 6

- Задача регрессии
- Линейные и нелинейные модели
- Практика на понимание основ линейной регрессии



Занятие 7

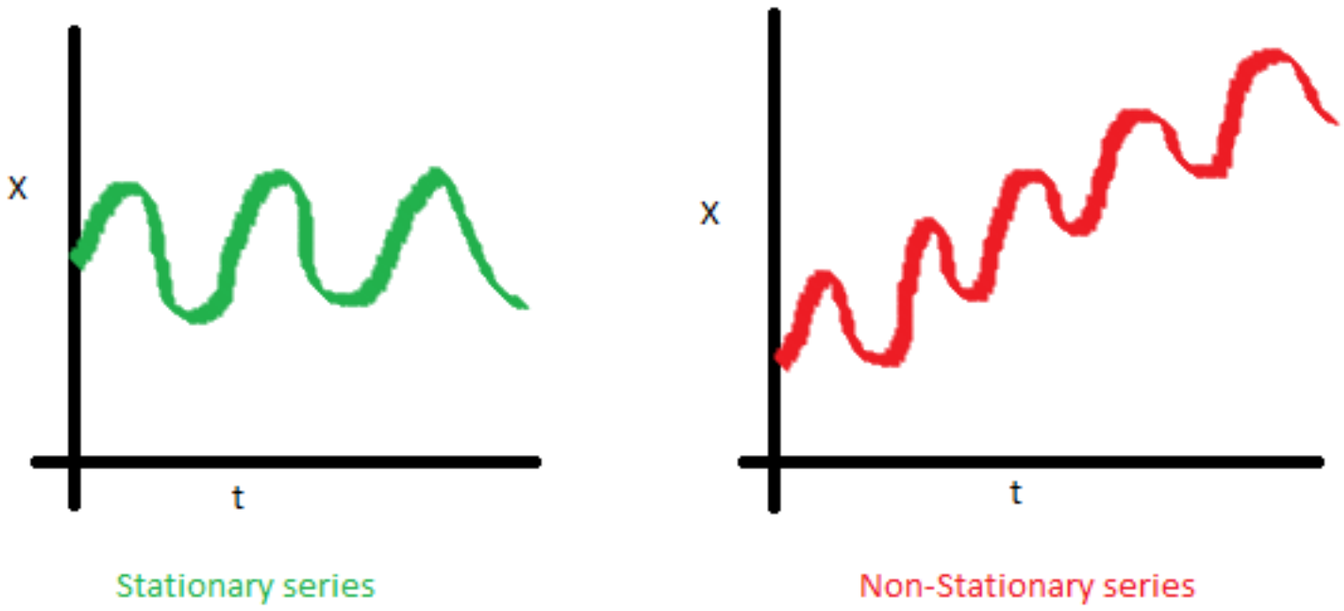
- Обучение без учителя
- Principal Component Analysis
- Кластеризация
- Практика на кластеризацию данных с Samsung Galaxy S3

Занятие 8

- Онлайн-обучение
- Обучение на гигабайтах
- Vowpal Wabbit
- Основы работы с текстами
- Практика на классификацию текстов по темам

Занятие 9

- Временные ряды
- Классика и современные подходы
- Лектор – Дмитрий Сергеев Zeptolab



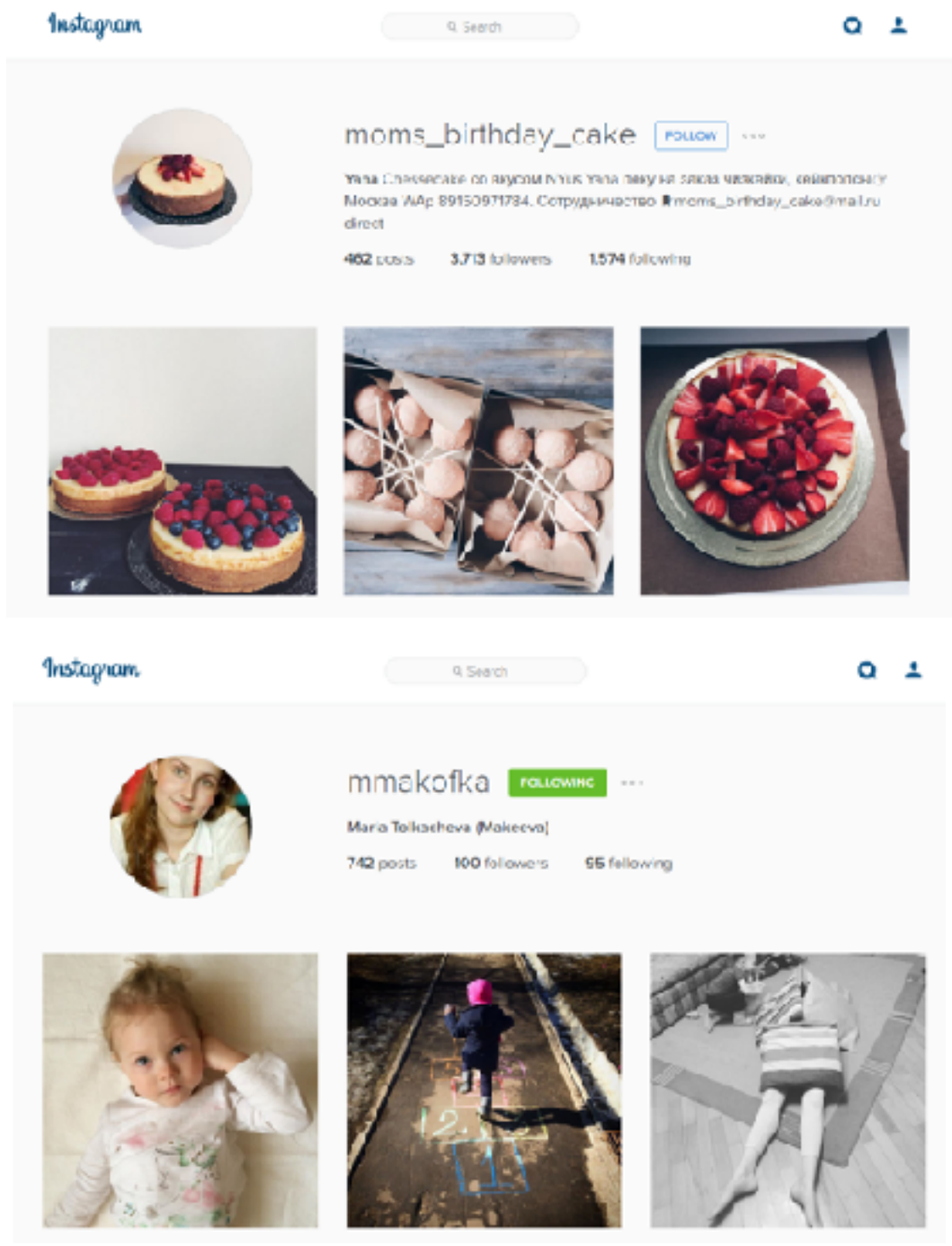
Занятие 10

- Современный взгляд на бустинг
- Теоретические основы
- Лучшие на сегодня реализации
- Лектор – Алексей Натекин, основатель ODS, ex. Deloitte

Перерегуляризировали

Индивидуальный проект

- В течение всего курса
- Четкий план
- Лучше свои данные
- Peer review
- Отличный опыт



Больше историй – в Slack

Open Data Science

Удачи!