



Машинное обучение и работа с большими данными

#1. Введение в анализ данных.



ПЛАН

#1. Введение в анализ данных.

01

ЗНАКОМСТВО
5 МИН



02

ВВЕДЕНИЕ
25 МИН



03

ОРГАНИЗАЦИОНКА
5 МИН



04

ВОПРОСЫ
5 МИН





Владислав Чертенков

26 годиков



НИУ ВШЭ | Аспирант 3 г.о.



Sber | ML Team Lead

tg: satankov

vchertenkov@hse.ru

Образование

2014 - Бакалавриат

№100 универ

2018 - Магистратура

МИЭМ НИУ ВШЭ

2020 - Аспирантура

МИЭМ НИУ ВШЭ

Мат. моделирование

Программирование

Python

Машинное обучение

PyTorch, Rapids, PL

Анализ данных

Статистика, тервер,
визуализация данных.



IT КОМПАНИИ

Строят цифровой бизнес



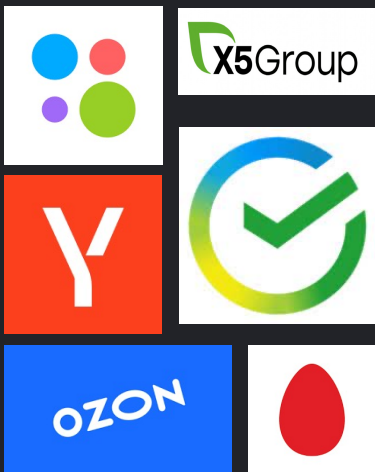
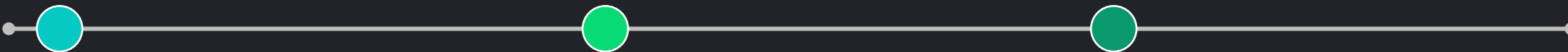
AGILE

Эффективный способ
организации команд



МЕТРИКИ

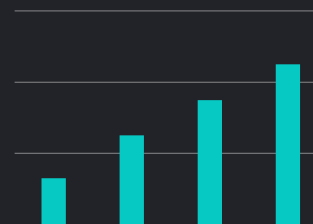
Рост доходов,
удовлетворенности клиентов



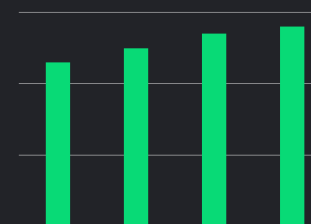
Состав команды:

- Менеджер
- Аналитик данных
- Разработчик
- Инженер
- Дизайнер

Money



Satisfaction



Делают это правильно с
помощью **математики**

Сегментация

- Выделение сегментов
- Оценка размеров аудитории
- Сравнение двух групп (сегментов)



Принятие решений

- Сравнение двух версий
- Выбор лучшего решения



Дизайн метрик

- Оценка качества
- Поиск неисправностей



Прогнозирование*

- Закупки
- Доходы, расходы
- Поломки



Другое

- Автоматизация
- Визуализация
- Поиск аномалий
- Генерация идей и поиск инсайтов

...





Выводы на
основе данных

Data analyst

Визуализация

Сегментация

Автоматизация

A/B тестинг

Дизайн метрик

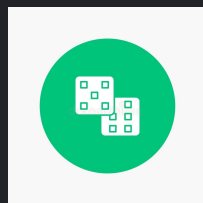
Прогнозирование

Методы
машинного
обучения

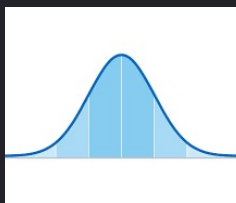
Нейронные
сети



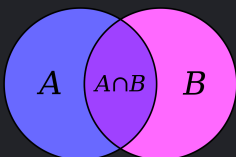
Data scientist



тервер



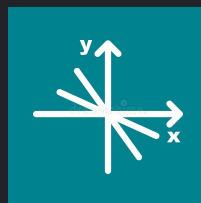
статистика



дискретная
математика



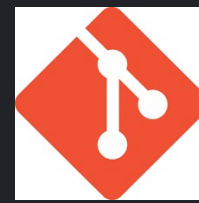
фурье анализ



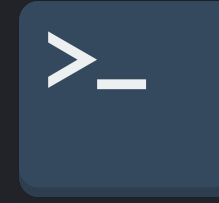
лин.алгебра



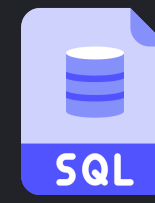
python + jupyter
notebook



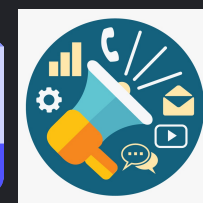
git



terminal



sql

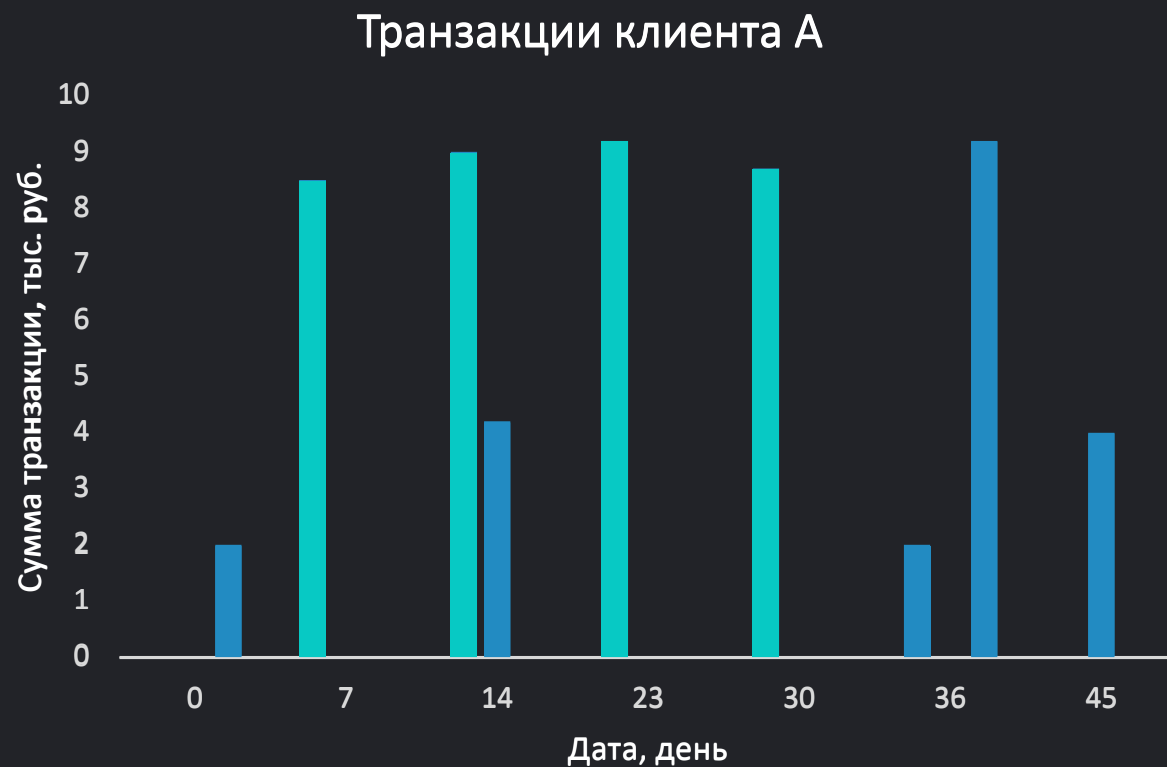


internet,
marketing



Задача

Научиться прогнозировать следующее действие, если человек совершает какое-то действие **регулярно**.

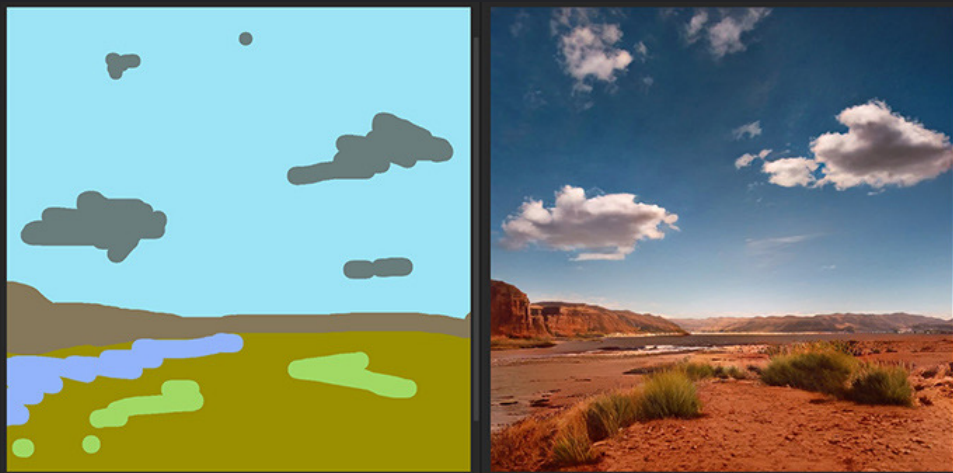
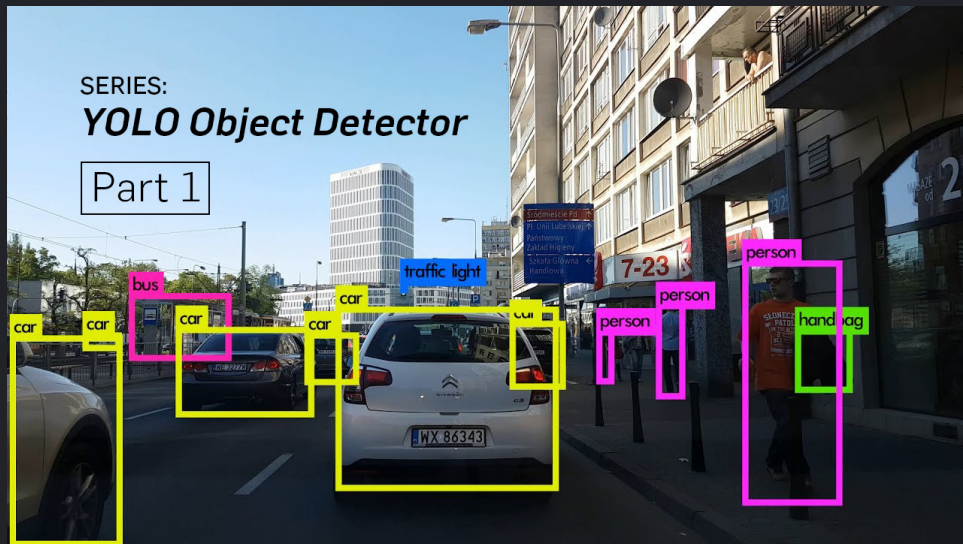


● Найдена 1 периодичность:

Сумма:
 8.9 ± 0.2 тыс. руб

Период:
 6.7 ± 0.5 дней

Когда:
2021-02-17 – 2021-03-15



Автопилот

- Сегментация
- Детекция объектов
- Распознавание знаков



3-D камеры

Оценка глубины без
применения лидаров с
помощью Computer Vision



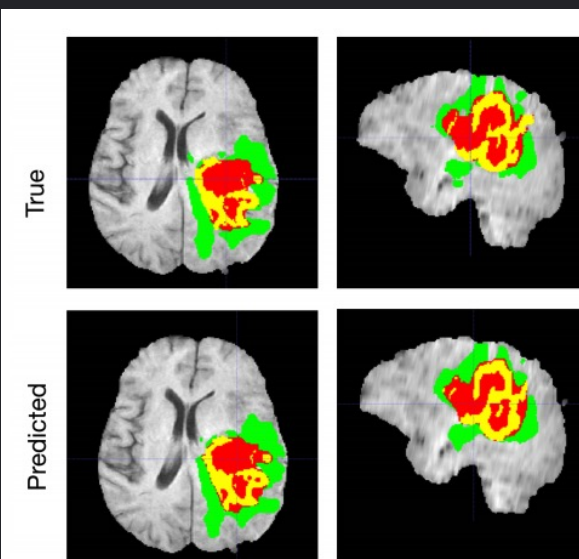
GAN

Генеративные сети: сами
рисуют и создают
объекты, которых не
существует в реальном
мире



Медицина

Помощь в обработке
изображений для
определения опухолей и
новообразований



03

ОРГАНИЗАЦИОНКА
5 МИН

Сдача домашки

В каждой домашке есть вопросы и ссылка на гугл-опрос. Заполняете результаты в гугл-опросе, я проверяю (автоматически). Вся подробная инфа будет в Jupyter Notebook`ах с домашками.

Практика и домашки

Практика и домашки выполняются в Jupyter Notebook. Установить Anaconda, библиотеки numpy, pandas, sklearn. Либо можно работать в Google Colab.



03

ОРГАНИЗАЦИОНКА

5 МИН

Вопросы

Все вопросы связанные с курсом задавать **только** в общем чате «**2023 Профсеминар**» в телеге, отметив меня @satankov. В личку с вопросами по поводу курса не писать. Все материалы: презентации, ноутбуки и задания буду выкладывать в телеге с тегом курса #машинное_обучение



Оценка

Оценка за курс где-то у вас в программе должна быть. Я точную формулу **не знаю**. Процент за мою часть курсаставляю за решение всех домашек в равной пропорции.