

академия  
больших  
данных



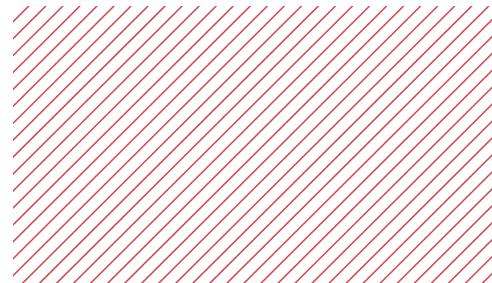
mail.ru  
group

# Поиск фрода в СитиМобил

Автор идеи: Алексей Чернобровов  
Ментор: Алексей Чернобровов

Команда:

Дмитрий Черников  
Олег Черемисин  
Алексей Мышлянов  
Константин Елизаров  
Павел Шальнов



# Описание проекта



# Описание проблемы/задачи

---

Первая часть: это исследование существующей литературы по Fake-GPS и поисках аномалий, а также других источников данных (например, сейчас мы дообогащаем наши данные данными вышек и вайфай сетей, но не всегда это хорошо работает со всеми устройствами).

Многие источники и методы уже устарели (как, например, развенчано поголовное использование цифровых отпечатков устройств, *fingerprint*), о которых мы еще не знаем или не подумали. Поэтому очень важно собрать наиболее широкий контекст методов решения вашей задачи, даже если эти решения не применимы к конкретному дата сету.

Вторая часть - это ML-модель, которая на основе данных находит подозрительные поездки, а также делает скоринг водителей на вероятность фрода.

Дополнительно можно построить модель, которая находит примеры, которые нужно доразметить, чтобы основная модель училась быстрее (*active learning*).



# Бизнес-задача

---

Выявление мошенничества (fraud detection) при использовании программных средств подмены геолокации (fake GPS)

- a. Где применять?

Агрегаторы и службы такси

- b. Кто пользователи?

b2b

- c. Аналогичные решения на рынке?

Jump Taxi (<https://blog.jump.taxi/fraudtaxidrivers>)

Такси Везёт (<https://habr.com/ru/post/512752/>)

<https://www.protaxi.su/all/taxi-frod/>



# ML-проблематика

---

Решение задачи semisupervised learning для выявления аномалий по данным геопозиционирования и показаниям акселерометра

*Задача может быть сформулирована как задача двоичной классификации (реальных и фейковые записи в логах)*



# Критерии успешности

---

Как мы будем оценивать качество модели:

у нас есть скоринг по аномальности поездок на различных метриках, таким образом, мы знаем на других данных на сколько водитель подозителен, качество модели вашей - это сумма скоринга водителей, которых вы разметите как фрод.



# На каждый пункт по слайду

---

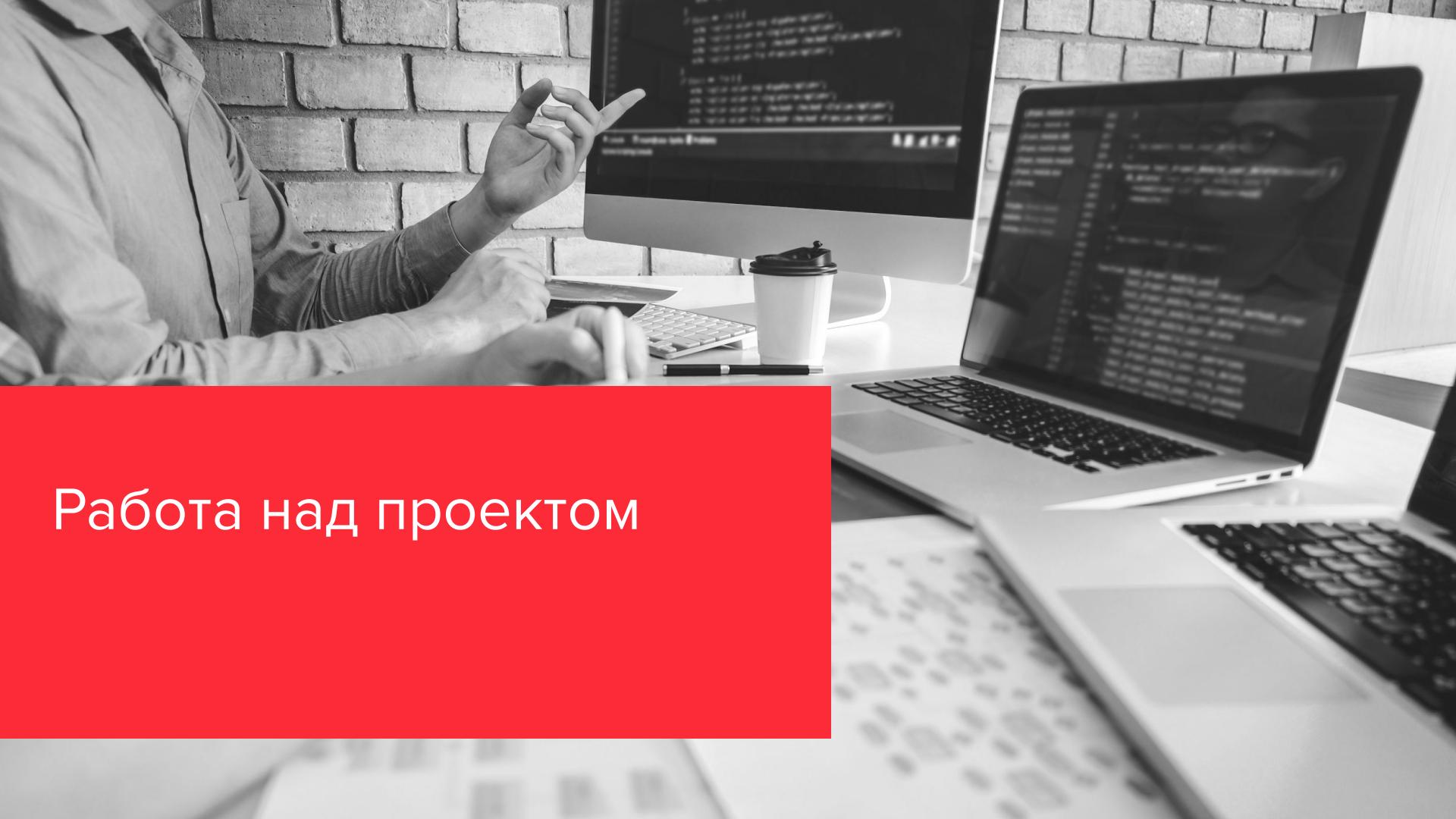
3. Что вы используете для реализации проекта?
  - a. Перечень инструментов и технологий, систем и т.п. техническая составляющая проекта (мелко базовый стек и крупно - что-то специфичное для проекта)
  - b. Данные. Открытые - какая лицензия / свои - как собирали, примеры
  - c. Выбор моделей. Обоснование, метрики качества, особенности обучения (если применимо), оригинальность решения (новые модели/приемы)
4. Оригинальность продукта - почему комбинация выбранных технологий может быть полезна, конкретный пример использования / демонстрации
5. Возможность решения существующих недоработок проекта (с примерами)
6. Интеграция в существующие сервисы / связка с тем, как/кто должен оценивать успех идеи (клиенты/бизнес/технологи)

Помните, вас придут слушать люди, которые воплощают идеи в бизнес-модели!

# Демонстрация проекта

262

263



Работа над проектом

## Product Roadmap



Hide

Roadmap Planning Board Parking Lot

Q1 2018

Jan

Feb

Mar

Q2

Apr

May

Q3

Jul

A

**Company All Hands**  
May 27, 2018

## Web Team

3rd Party Integrations

Ticketing System

New Admin Console

Self Service Portal

Security 2.0

On Premise Backup

Code Review

Shopping Cart Improvements

API

## Mobile Team

**Release**  
Mar 24, 2018

Mobile Mock-Up

UX Improvements

Cloud Support

UX Improvements

Application Upgrade

Android Application

Interactive Dialogue Box

Automatic Renewal Service

Clean up CSS on user portal

Q3 Initiative

## Marketing Team

Market Analysis

Customer Outreach

SEO Plan

Settings

Filter

Share

Version

Export

History

Add Bar

Add Container

Add Milestone

Add Lane

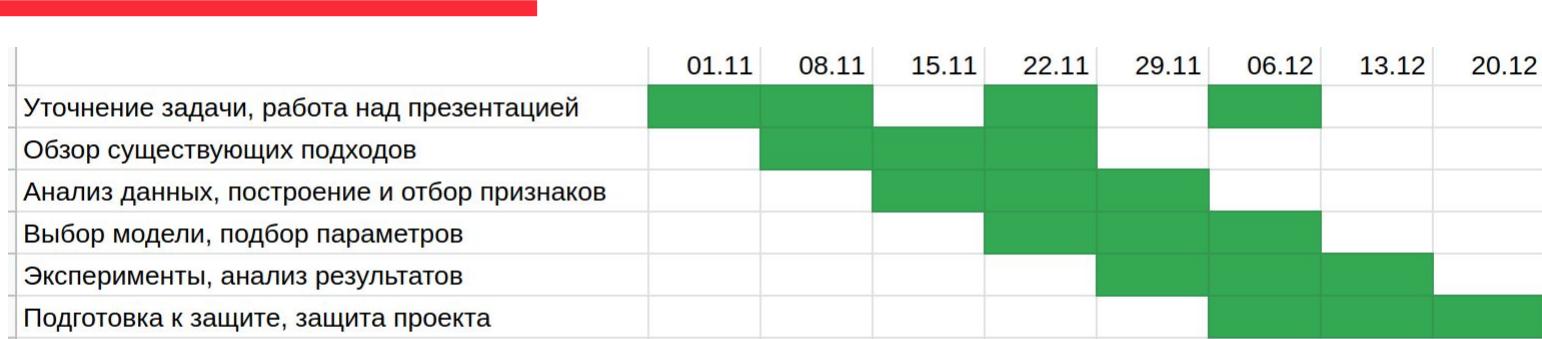
**Goals**

Enhance Performance

Increase Customer Satisfaction

Increase Revenue

# Roadmap





# Команда

---

Дмитрий Черников - распределение задач

Олег Черемисин - распределение задач

Алексей Мышлянов - распределение задач

Константин Елизаров - распределение задач

Павел Шальнов - распределение задач