Практический анализ данных и машинное обучение

Факультет Компьютерных Наук ВШЭ | Виталий Шайдуров

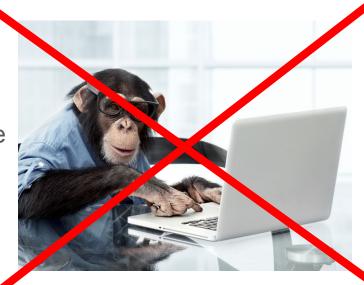
Что нас ждет

- 11 занятий
- Основные алгоритмы и их использование
- Домашние задания и практики
- Соревнование Kaggle Inclass
- Индивидуальные проекты



Особенности курса

- Обилие практики задания на каждом занятии и после него
- Четкие инструкции к заданиям (макет в виде тетрадки в Jupyter Notebook)
- Понимание теоретических основ алгоритмов
- Знакомство с платформой Kaggle
- В основе всего -- свой собственный проект



Организационные вопросы

- Общение -- в Telegram
- За домашние задания тах 10 баллов
- За проект и соревнование тах 30 баллов
- Текущий рейтинг будет <u>тут</u>
- Нужны аккаунты GitHub и Kaggle
- Все материалы курса -- в проекте на <u>GitHub</u>

Prerequisits

- Минимальное владение Python
- Основы аналитической геометрии (вектора, матрицы, скалярное произведение и т.д.)
- Основы математического анализа (производные, пределы, интегралы)
- Теория вероятностей и статистика (этого будет не много)

Инструменты

- Язык Python
- Jupyter notebooks
- Сборка библиотек Anaconda
- GitHub
- Kaggle
- XGBoost и др.











Программа. Часть І

- Основы машинного обучения
 - Обучение с учителем. Задачи классификации и регрессии
 - Оценка качества алгоритмов машинного обучения
 - Обучение без учителя и задача кластеризации
 - Поиск выбросов и аномалий в данных

Программа. Часть II

- Продвинутые методы машинного обучения
 - Ансамбли алгоритмов классификации и регрессии.
 Градиентный бустинг
 - Смешивание моделей классификации и регрессии. Стекинг моделей классификации и регрессии
 - Соревнования по анализу данных. Обзор решений

Индивидуальный проект

- В течение всего курса
- Лучше свои данные
- Подробный отчет по проекту
- Презентации в конце

Занятие 1

- Анализ данных с NumPy, SciPy, Pandas
- Практика на знакомство с данными
- Д3 № 1. Анализ демографических данных по жителям США