

Открытие Школы-ИТ Data Science начнется
в 19:00

План занятия

- Знакомство, орг вопросы
- Разбор домашнего задания
- Разведочный анализ данных (EDA)
- Разбор семинарского ноутбуку по EDA

Давайте знакомиться!

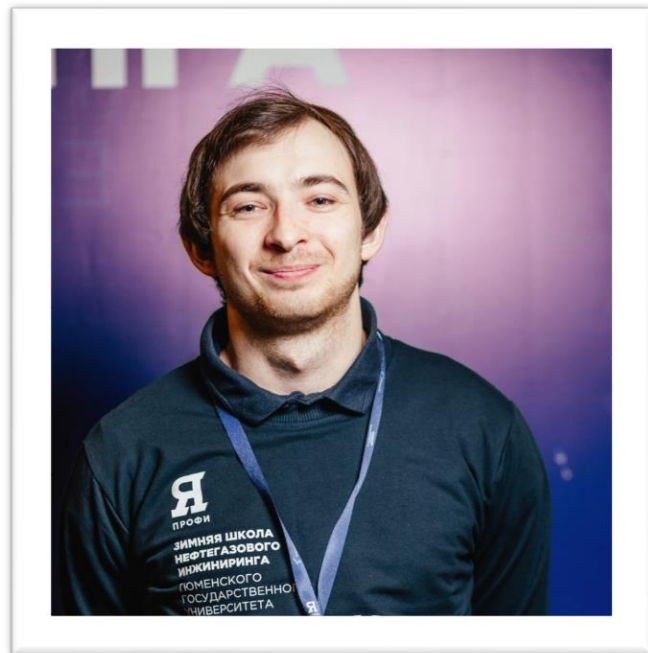
Преподаватель

Ринат Плавник

Закончил бакалавриат и магистратуру МФТИ,
диплом с отличием

Опыт работы:

- Нефтегазовый центр МФТИ
- Лаборатория моделирования механических систем
и процессов



tg – @PlavRin

О курсе

Правила присутствия на занятиях

- В Zoom нужно заполнить **реальные Фамилию и Имя**
- У всех участников **выключен звук** во время выступления спикера
- У всех участников **включены камеры**
- Обязательно нужно **задавать вопросы** в чат, если что-то непонятно

Не стесняйтесь перебивать меня
и задавать вопросы :)

Чем интересен DS

Работа специалиста по анализу данных – это работа детективом в области данных.

В большинстве университетов ставится только вопрос «как решить задачу», без вопроса «зачем» – это демотивирует!

Мы будем от интуиции и постановки проблемы двигаться в сторону математических понятий. Математика – это красиво, когда ты понимаешь зачем это 😊

Будет сложно, но вам понравится!

Зачем нужен курс

- Получить хорошую базу для освоения профессии в сфере Data Science, чтобы стать хорошим и востребованным специалистом
- Оказаться в среде таких же заинтересованных людей
- Научиться применять математику и Python для решения бизнес-задач

Коммуникация между командой и слушателями

- Не бойтесь задавать вопросы на занятиях
- Вы будете получать оперативную обратную связь на решения домашних заданий
- Задавайте любые вопросы в чате (мы с радостью ответим!)

Что делать, если у меня возник вопрос?

- Ищем в google или на [stackoverflow](https://stackoverflow.com) (Крайне важно научиться это делать!)
- Спрашиваем в общем чате у однокурсников
- Спрашиваем у преподавателей и ассистентов

В случае форс-мажорных обстоятельств нужно сразу писать Куратору

Расписание занятий

- 3 недели занятий
- 1 неделя защита командного проекта EDA и Каникулы
- 3 недели занятий
- 1 неделя защита командного проекта ML и Каникулы
- 3 недели занятий
- 3 недели на защиту финального проекта

Расписание занятий:

- Среда, 19:00-22:00 мск
- Суббота, 10:00-13:00 мск

Программа курса

Первичный анализ данных в Python

1 неделя: Закрепление материалов по первичному анализу

Классическое машинное обучение

2 неделя: Введение в машинное обучение. Математика для машинного обучения

3 неделя: Задачи регрессии и классификации. Защита проекта по первичному анализу данных

Алгоритмы и прикладные задачи классического ML

5 неделя: Углубленная математика и линейные модели

6 неделя: Деревянные алгоритмы

7 неделя: Повторение ML. Задача временных рядов

Основы Deep Learning

9 неделя: ООП и работа с картинками как с матрицами. Введение глубокое обучение.

10 неделя: Задачи компьютерного зрения

11 неделя: Задачи NLP

Выпускной проект

12 неделя: Консультация

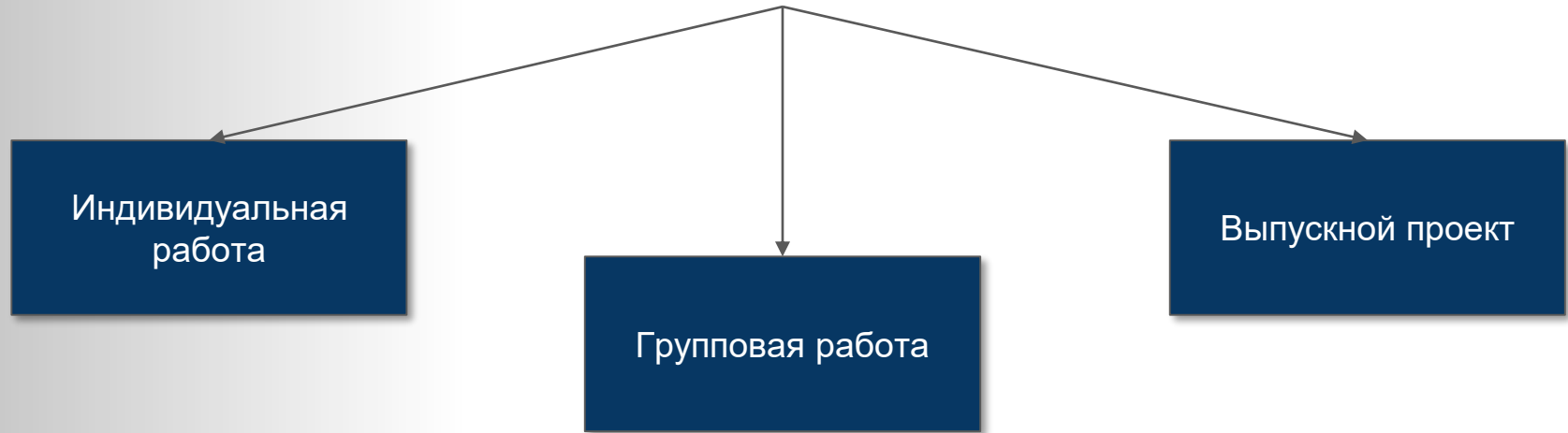
13 неделя: Предзащита

14 неделя: Финальная защита

Разумеется, это далеко не всё,
что вам нужно будет знать...

...но этого достаточно, чтобы дать
каждому из вас определиться с
траекторией своего развития

Домашние задания



Домашние задания: индивидуальная работа

Мини дз:

- *Короткие и простые*
- *Даются ~после каждого занятия*
- *Дедлайн: 3 дня после выдачи*

Сложные дз:

- *Более интересные*
- *Даются ~раз в одну - две недели*
- *Soft deadline: 1 неделя, hard deadline 2 недели. В течение второй недели штраф 50%, а после работа не будет оцениваться.*

Формат сдачи:

- Сдача заданий в Google Classroom
- Проверка в течение 3 дней после дедлайна
- Получение фидбека от проверяющего ассистента
- Доделывание и исправление задания в случае замечаний

Каждое задание будет оцениваться в диапазоне от 0 до 1, где 1 – полностью правильно выполненная работа

Домашние задания: групповая работа

Два групповых проекта:

- Проект “Первичный анализ данных”
- Проект “Применение методов ML”

В случае изменения сроков защиты проектов вас заранее оповестят

Формат сдачи:

- Работа в команде (информация про распределение будет доступна позже, но можно собраться самим по 3-4 чел.)
- Публичная защита на дополнительных занятиях
- Подробная информация будет доступна в Google Classroom

Выпускной проект

Виды проектов:

- Тема по рекомендации преподавателя
- Индивидуальная тема

Порядок действий:

1. Выбрать тему после первой половины курса
2. Защитить выбранную тему
у преподавателя
3. 13 неделя: консультация и предзащита
4. 14 неделя: консультация и защита
5. Успешная сдача!



Примеры выпускных проектов

Примеры проектов из прошлых запусков:

- “Автоматизация таргетирования рекламы сервиса оплаты по QR-коду среди частных предпринимателей”

Бизнес-аналитика

- “Система автоматического детектирования поломок серверных стоек на основе данных с камер наблюдения”

Компьютерное зрение

- “Прогноз времени рассмотрения заявки на проведение залоговой экспертизы” *Бизнес-аналитика + NLP*



Система оценивания

Для получения зачета за профильный модуль необходимо:

- выполнить хотя бы один групповой проект
- сдать ≥ 9 мини-ДЗ
- сдать ≥ 4 больших ДЗ

Итог – защита выпускного проекта. За проект можно получить от УД до ОТЛ.

Таблица оценок будет выложена в канал в течение недели

Дальнейшие действия

Что дальше

- Вступить в чат и подписаться на канал в Telegram
- Выполнить нулевое домашнее задание

Домашнее задание 0 (дедлайн – 17.12 23:59)

1. Завести Google-аккаунт и добавиться в наш Google Classroom (указать реальное полное имя и сделать это только один раз!).

Почта должна оканчиваться на @gmail.com.

2. Завести аккаунт на [LeetCode](#)

Для того, чтобы быть DS-специалистов уровням топ-компаний РФ, вам, увы, придётся знать алгоритмы хотя бы на уровне LeetCode-Easy.

Начните практиковаться заранее.

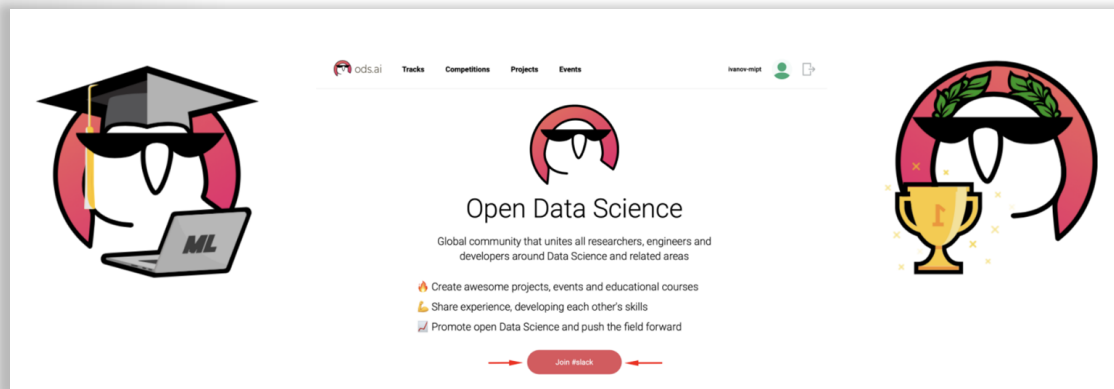
3. Завести аккаунт на [Kaggle](#)

Kaggle — лучшая платформа для того, чтобы оттачивать навык решения DS-задач. Решение нескольких домашних заданий будет проходить там.

Домашнее задание 0 (дедлайн - 17.12 23:59)

4. Вступить в сообщество Open Data Science

Data Science развивается слишком быстро. Чтобы успевать за трендами, вам нужно стать частью сообщества. Open Data Science <https://ods.ai/>



Домашнее задание 0 (дедлайн - 17.12 23:59)

5. Data analyst, data scientist, data engineer... Кто чем на самом деле занимается?

В области анализа данных есть много специализаций с похожими названиями и пересекающимися обязанностями.

Чтобы окончательно разобраться в том, чем именно занимается data scientist, послушайте выступление

Алексея Намёкина, основателя ODS. <https://www.youtube.com/watch?v=IDkTNURDlaY>

6. Заполнить [#whois](#) в чате группы

Обратная связь – обязательная часть занятий!

После каждого занятия заполняем
форму обратной связи:

https://docs.google.com/forms/d/e/1FAIpQLSfmBbkDq3WQVo43ro5k9A/viewform?usp=sf_link



До встречи на следующих занятиях!