

Программа курса

«MLOPS: Разработка и внедрение ML-решений»

О курсе: 6-дневный практический курс MLOps — это комплексная программа обучения для ML инженеров, которая проведет вас через все этапы жизненного цикла ML-проекта: от сбора и подготовки данных до развертывания, мониторинга и автоматизации моделей в production.

Аудитория:

- Data Scientist
- ML инженер
- Разработчик ПО
- DevOps-инженер

Уровень подготовки:

- Опыт программирования на Python
- Основы анализа данных

Продолжительность курса: 24/40 академических часов, 6/10 дней по 4 ак. часов дистанционно

Содержание программы

1. Жизненный цикл разработки ML-модели

- **Основные этапы жизненного цикла ML-модели:** постановка задачи, подготовка данных, выбор модели, обучение и оценка.
- Ключевые типы задач (классификация и регрессия), метрик качества, а также архитектура ML-сервисов.
 - **Практическая часть:** Развертывание среды разработки с использованием шаблона проекта. Реализация задачи классификации и регрессии на примере предсказания стоимости автомобиля.

2. Управление кодом, данными и признаками

- Подходы к командной разработке ML-решений: работа с Git, линтерами, code review.
- Знакомство с DVC для версионирования данных и Feature Store (Feast) как хранилищем признаков.
- Разбор лучших практик организации и воспроизводимости экспериментов.
 - **Практическая часть:** Практика использования Git и DVC в ML-проекте (в качестве удаленного хранилища данных – S3). Освоение Feast для хранения и запроса фичей.

3. Управление экспериментами и артефактами

- Инструменты логирования и визуализации экспериментов.
- MLFlow и ClearML как платформы для управления метриками, параметрами, моделями и артефактами.
- Хранение моделей и данных на внешних хранилищах (например, S3).
 - **Практическая часть:** Настройка и использование MLFlow и ClearML. Сохранение параметров и метрик, экспериментов, сохранение моделей и других артефактов. Интеграция с S3-хранилищем.

4. Развёртывание моделей и CI/CD

- Подходы к CI/CD в ML-проектах.
- Обсуждение юнит-тестирования, автоматизацию обучения и развёртывания.
- Варианты серверов для инференса: Triton Inference Server.
 - **Практическая часть:** Создание сценария инференса моделей.

5. Оркестрация ML-пайплайнов с Apache Airflow

- Архитектура и возможности Apache Airflow.
- Роль оркестрации в ML: планирование, зависимость задач, мониторинг процессов.
 - **Практическая часть:** Создание DAG в Airflow. Организация пайплайна обработки данных и инференса модели.

6. Мониторинг моделей и визуализация

- Мониторинг моделей и данных в проде.
- Подходы к отслеживанию дрейфа, качества и стабильности работы моделей.
- Знакомство с инструментами Superset и Evidently.
 - **Практическая часть:** Создание дэшбордов в Superset. Настройка мониторинга модели через Evidently: отслеживание дрейфа и метрики.

Итоговый проект

Ваш индивидуальный ML проект — это ваша возможность продемонстрировать все, чему вы научились на курсе. на протяжении всего курса Вы сможете реализовать собственный проект или предложенный нашими преподавателями. В рамках проекта вы пройдете все этапы MLOps-цикла:

- Постановка задачи и подготовка данных.
- Обучение и версионирование модели.
- Управление экспериментами.
- Развертывание модели как сервиса.
- Настройка CI/CD пайплайна.
- Настройка мониторинга и алертинга.