

# **О курсе MLOps**

Павел Кикин - Head of MLOps

Владислав Горбунов - Head of DS

Ирина Беккер - Куратор курса

# Проблемы, которые нужно решать

- Нужно вместе с коллегами сделать общее решение с ML-моделями
- Нужно принять выполненные задачи от коллег
- Нужно принять результаты работы подрядчиков и продолжить выполнение проекта
- Нужно подключиться в другой проект с кучей legacy потому что кто-то заболел / ушел в отпуск / уволился
- Нужно поставить решение в прод и передать на сервисную поддержку
- Нужно воспроизводить пайплайны и пере-обучать модели для поддержки решений в проде
- Нужно подготовить материалы по проекту для рассказа inhouse или на конференции

Какие темы покрывают данные проблемы:

- Работа в команде над DS / ML проектами
- Воспроизводимость DS исследований
- Вывод моделей в продуктив, автоматизация и поддержка

# Фокус внимания на двух составляющих работы

- **Исследования в DS / ML** сосредоточены на изучении и анализе данных для получения новых знаний, а также экспериментировании, разработке и тестировании новых алгоритмов и моделей. Разработанная модель - только часть большой системы, сама модель часто ценности не несет
- **Продуктивизация в DS / ML** фокусируется на интеграции построенных моделей в бизнес-процессы, масштабирование их использования, обеспечение стабильности и поддержки в рабочей среде (деньги находятся тут)

# Что Вам даст этот курс?

- Как работать в команде над одним проектом и организовать процесс ML разработки в команде
- Превратить модель в сервис
- Корректно передавать результаты исследований заказчику или другим разработчикам
- Систематизировать исследования
- Отслеживать и сохранять условия и результаты экспериментов
- Автоматизировать процесс проведения стандартизированных исследований
- Повысить качество кода
- Обеспечить воспроизводимость ваших исследований

# КТО МЫ?

## Roles:

- Head of MLOps
- Head of DS
- TechLead DS
- Senior DS's
- Senior MLE's

## Spheres:

- Oil and Gas,
- Banking,
- Social Network,
- EdTech,
- Education

# Организационные моменты:

- Как начать участие?
- Какие занятия ждут на курсе?
- Домашние задания
- Лидерборд
- Прохождение курса в команде и самостоятельно
- Дополнительные активности
- Итоговые проекты

# Критерии оценки проектов

- Цель проекта; его ценность; бизнес-метрика; связь с техническими метриками; список гипотез, которые проверяли; почему проверяли именно эти гипотезы;
- Организация рабочего процесса (процессы работы команды, приоритезация списка гипотез, рабочее место dsa)
- Использование контроля версий кода и инструментов командной разработки (git, code review, linters, formatters, CI)
- Чтение и слияние исходных данных из различных файлов/форматов (в т. ч. из разных источников, с выгрузкой в версионированное хранилище)
- Исследования на данных (EDA - графики по имеющимся значениям, проверка распределения данных/стационарность, расчет стандартных статистических характеристик; подготовка интерактивных отчетов по данным и гипотезам);
- Предобработка данных (исправление некорректных значений и выбросов, фильтрация, замена пропусков и т.д., сохранение и версионирование датасетов, автоматизация пайплайнов подготовки датасетов);
- Реализация пайплайнов обучения и валидации моделей, проведение сравнительного анализа моделей (сохранение и версионирование моделей и пайплайнов, бэнчмарк моделей: время работы, потребление ресурсов)
- Реализация сервиса с интерфейсом (интерфейс, масштабирование, очереди, мониторинг)
- Сборка и поставка сервиса с моделью (CI/CD, поставка моделей)
- Бонус от организаторов курса (бонусные критерии скрыты для избежания хакинга метрики)

# Доклады - примеры тем

- Исследование сервисов управления ноутбуками (sagemaker, databricks, datalore, ванильный JH)
- СУБД
- Особенности ООП в МЛ
- W&B
- neptuna
- Pachyderm
- Dataflow
- ONNX, PMML
- Improving Prediction Latency and Reducing Resource Costs
- Ваши темы...

# Доклады - как взять на подготовку?

1. Обратитесь к [@Irinka\\_Bekker](#) с предложением своей темы (тема должна соответствовать курсу, мы будем встраивать доклад в соответствующий раздел, чтобы не терять логику повествования)
2. Получите ок на подготовку темы
3. Подготовьте материалы
4. Пройдите ревью материалов
5. Подготовьте выступление и пройдите прогон с менторами / авторами курса
6. Выступите онлайн или подготовьте запись
7. Получите бонусные баллы

**Удачи!**