

Инженер машинного обучения

01	Кому подойдёт курс	Разработчикам и инженерам данных: - углубите знания по алгоритмам и инструментам машинного обучения, чтобы участвовать в сложных проектах и быть конкурентоспособным на рынке труда Аналитикам и специалистам по Data Science: - пройдёте через полный цикл ML-проектов: от разработки модели до её внедрения — чтобы выйти на новый уровень развития карьеры Начинающим ML-инженерам: - сможете структурировать знания, попрактиковаться, получить обратную связь от более опытных коллег и повысить свою квалификацию
02	Чему научитесь на курсе Какие знания и навыки освоите	За 4 месяца освоите полный цикл ML-проектов. Сможете сами: - Собирать данные и запускать с ними регулярные процессы - Обучать модели в стабильной инфраструктуре и выводить их в продакшен - Находить ошибки в процессах, улучшать базовый пайплайн и модели - Создавать и контейнеризировать микросервисы - Работать с uplift-моделированием - Строить модели неперсонализированных рекомендаций, контентной и коллаборативной фильтрации Разработаете 3 сервиса: - на данных Яндекс Недвижимости обучите модель, которая предсказывает события. Создадите микросервис и развернёте его в облачной инфраструктуре - разработаете рекомендательную систему для улучшения предсказаний в сервисе Яндекс Музыка - решите задачу uplift-моделирования: определите аудиторию и финансовый эффект
03	Как проходит курс	- Теория и практика на платформе Практикума - Практические задания на готовой инфраструктуре - Воркшопы - Индивидуальные онлайн-встречи с наставником - Доступ из любой точки мира в удобное время

Что вас ждёт

Обучение продвинутых моделей и создание микросервисов в облачной среде	6 ML-проектов в портфолио, включая настоящие сервисы	Docker, FastAPI, Airflow, MLflow, Yandex Cloud и другие инструменты
--	--	---

Инженер машинного обучения

4 месяца

продолжительность курса

6 ML-проектов в портфолио

включая настоящие сервисы

2 ЧАСА

00

Бесплатная часть.
Создание пайплайна
по подготовке данных
и выводу ML-модели
в продакшн

- Знакомство с курсом
- Кто такой ML-инженер
- Пайплайны подготовки данных в Airflow
- Создание baseline-модели
- Улучшение baseline-модели
- Релиз ML-модели
- ML-модель в продакшене
- Рекомендательная система и uplift-модель
- Входной тест

3 НЕДЕЛИ

01

Разработка пайплайнов
подготовки данных
и обучения модели

- Подготовка и трансформация данных в Airflow
- Создание ML-модели и настройка DVC-пайплайна

3 НЕДЕЛИ

02

Улучшение baseline-модели

- Контроль за экспериментами
- Конструирование признаков: Feature Engineering
- Подбор гиперпараметров модели

3 НЕДЕЛИ

03

Релиз модели в продакшен

- Создание микросервиса
- Контейнеризация микросервиса
- Мониторинг ML-модели

3 НЕДЕЛИ

04

Создание рекомендательной
системы

- Рекомендательные системы
- Базовые рекомендации
- Двухстадийный подход
- Архитектура рекомендательных систем

2 НЕДЕЛИ

05

Создание uplift-модели

Uplift-моделирование

3 НЕДЕЛИ

06

Итоговый проект

Полный цикл работы с ML-моделью



Факультативный курс.
Сферы применения
машинного обучения

Машинное обучение в бизнесе,
для текстов и с учителем

Бесплатная часть. Создание пайплайна по подготовке данных и выводу ML-модели в продакшн 00

~2 часа

Входное тестирование

- Познакомитесь с ролью и задачами ML-инженера, узнаете, в чём разница между ML-инженером и другими специалистами
- Узнаете, как использовать Airflow, чтобы автоматизировать процесс подготовки данных, что такое DAG, как создавать пайплайны, а также как обрабатывать данные на разных этапах
- Разберётесь, что такое baseline-модель, подберёте метрики для оценки качества модели и запишите результат в MLflow
- Улучшите baseline-модель, чтобы повысить качество и точность прогнозов
- Поймёте, что такое микросервис, зачем он нужен бизнесу, а также как создавать простой микросервис на фреймворке FastAPI
- Познакомитесь с библиотеками Python, которые позволяют отправлять HTTP-запросы к микросервису с моделью и получать результаты
- Разберётесь, что такое рекомендательная системы и uplift-моделирование, а также как они применяются в реальном бизнесе
- Пройдёте входное тестирование, проверите знания и решите, подходит ли вам курс

Практическая работа

Решите несколько задач из жизни ML-инженера, пройдёте входное тестирование и поможёте, подходит ли вам курс

Инструменты и технологии

- Airflow
- SQL
- MLflow
- Optuna
- Python
- FastAPI
- Yandex Cloud

Разработка пайплайнов подготовки данных и обучения модели

01

3 недели

5 тем

Проект

Научитесь подготавливать и трансформировать данные в Airflow: знакомство с Airflow, первый DAG, подготовка данных для обучения базовой модели.

Создайте ML-модель и настройте DVC-пайплайн: автоматизация подготовки данных (ETL) с помощью Airflow, обучение базовой модели, создание и запуск DVC-пайплайна, версионирование артефактов.

Проект

Для сервиса Яндекс Недвижимость настройте процесс обучения модели, которая должна помочь остановить отток клиентов

Инструменты и технологии

- DVC
- AirFlow
- S3

Улучшение baseline-модели

02

3 недели

6 тем

Проект

Изучите основные сущности MLflow: Tracking, Projects, Models, Model Registry, развёртывание MLflow, логирование ML-модели, сохранение окружения, визуализацию результатов, методы версионирования модели и контроль изменений, доступ к результатам и артефактам моделей через API.

Освойте конструирование признаков: Feature Engineering — продвинутый EDA (Exploratory Data Analysis), работу с разными типами признаков (ускорение обработки, применение автоматической генерации); отбор наиболее важных признаков, организацию обработки данных.

Научитесь подбирать гиперпараметры модели: стандартные методы подбора (решётчатый поиск (Grid Search) и случайный поиск (Random Search), байесовский подход); фреймворк Optuna, визуализацию процесса подбора гиперпараметров — Optuna и MLflow одновременно.

Проект

Улучшите качество модели с помощью Feature Engineering и оптимизации гиперпараметров

Инструменты и технологии

- MLflow
- Optuna
- Mlxtend
- Продвинутый EDA



3 недели
6 тем
Проект

Создадите микросервис: архитектуру приложения, использование FastAPI для написания микросервиса на базе ML-модели, его сборку, отладку и запуск.

Контейнеризируете микросервис: настройку изолированного окружения, использование готовых Docker-образов и запуск Python-скриптов, Docker-файл для запуска, оптимизацию разработки и запуска приложений — Docker Compose.

Настроите мониторинг ML-модели: метрики мониторинга в продакшене, их сбор, хранение (Prometheus Server), аналитику (PromQL) и визуализацию в Grafana.

Проект

Создадите сервис на базе ML-модели и развернёте его в облачной инфраструктуре

Инструменты и технологии

- Docker
- FastAPI
- Yandex.Cloud
- Uvicorn
- Prometheus
- Grafana

Создание рекомендательной системы

04

3 недели

7 тем

Проект

Познакомитесь с рекомендательными системами:

- принципы и алгоритмы персональных рекомендаций, виды рекомендательных систем, «холодный старт», матрица взаимодействий;
- базовые рекомендации — коллаборативная фильтрация (ALS), контентные рекомендации, валидация рекомендательной системы;
- двухстадийный подход — генерация наборов рекомендаций с помощью разных алгоритмов, ранжирование полученных рекомендаций для отбора лучших из них;
- архитектура рекомендательных систем: типовая архитектура, онлайн-рекомендации, сервис рекомендаций: ASGI-сервер Uvicorn на базе фреймворка FastAPI.

Проект

Разработаете рекомендательную систему для улучшения предсказаний в Яндекс Музыке, оформите сервис и выкатите его в продакшен

Инструменты и технологии

- Uvicorn
- implicit
- catboost
- U2U, I2I, ALS
- S3

Создание uplift-модели

05

2 недели

5 тем

Проект

Познакомитесь с uplift-моделированием: концепцией, метриками и методами оценки качества, базовыми методами, продвинутыми подходами, продакшеном.

Проект

Решите задачу uplift-моделирования: определитесь с аудиторией и оцените финансовый эффект

Инструменты и технологии

- S/T-learner
- scikit-uplift
- X/R-learner
- uplift-деревья

3 недели
Проект

Подтвердите, что освоили новые навыки, и пройдёте полный цикл работы с моделью: от формирования задачи и сбора данных до разработки модели и выкатки её в продакшен.

Факультативный курс. Сферы применения машинного обучения

- Научитесь с помощью ML-моделей запускать новые фичи и обрабатывать текстовые данные, а ещё поймёте, как упростить обучение моделей.
- Машинное обучение в бизнесе: запуск новых функций сервиса с применением машинного обучения, сбор данных, связь метрик машинного обучения с метриками продукта.
- Машинное обучение для текстов: классификация для текстовых данных, векторные представления текстов с помощью статистических моделей, работа с языковыми представлениями нейросетей.
- Машинное обучение с учителем: новые модели машинного обучения, метрики, средства для улучшения моделей и данных, упрощение обучения моделей.
- Инструменты и технологии**
- ML и бизнес
 - Метрики машинного обучения
 - Векторные и языковые представления
 - Улучшение моделей