

ML-инженер

01 Кому подойдёт курс

Учиться будет несложно, если вы:

- **Новичок в ML с технической базой** — знаете синтаксис Python и можете выполнять базовые операции
- **Математик без опыта разработки на Python** — понимаете линейную алгебру, теорию вероятностей и статистику
- **Аналитик без опыта в ML** — можете писать базовые SQL-запросы и знаете основы Python

02 Чему научитесь на курсе

- Строить и оптимизировать ML-модели
- Создавать автоматизированные ML-пайплайны
- Разрабатывать веб-сервисы на FastAPI и контейнеризовать их для продакшена
- Настраивать A/B-тесты и мониторинг качества модели
- Вести эксперименты и версионирование моделей в MLflow

03 Как проходит обучение

- Теория и практика в тренажёре на платформе Практикума
- Индивидуальные и групповые проекты с проверкой ревьюерами на данных из разных индустрий
- Регулярные воркшопы с экспертами
- Общение с одноклассниками и наставниками в чате курса

Что вас ждёт на обучении

Практика, основанная на решении реальных рабочих задач

Обучение от экспертов из Яндекса и других крупных компаний

Подготовка к техническим собеседованиям и презентация прошлого опыта

ML-инженер

Сравнение тарифов

	ML-инженер	ML-инженер расширенный
Длительность	12 месяцев	16 месяцев
Нагрузка	15 часов в неделю	15 часов в неделю
Уровень знаний на выходе	Junior	Junior
Проекты в портфолио	18 проектов	25 проектов
Проекты в Мастерской	По желанию, после окончания курса	Обязательно
Помощь в трудоустройстве	С середины обучения	С середины обучения
Недоступна при оплате курса работодателем		
Подготовка к техническим собеседованиям	В рамках основного курса	В рамках основного курса
Недоступна при оплате курса работодателем		
Каникулы	5 раз	5 раз
Продолжительность одних каникул	Неделя	Неделя
Сопровождение: наставник, ревьюеры, куратор, техническая поддержка	✓	✓
Воркшопы с экспертами	1 раз в 2 недели	1 раз в 2 недели
Доступ к выпускному сообществу	✓	✓
Диплом о профессиональной переподготовке для студентов с дипломами ВО/СПО.	✓	✓
Для остальных — сертификат об окончании курса		

ML-инженер

Преимущества расширенного тарифа

01	Два дополнительных модуля: Работа с большими данными, Обработка текстов и LLM	<p>В расширенный тариф входит изучение двух дополнительных доменных специализаций, которые усиливают ваши возможности трудоустройства: работа с большими данными и NLP/LLM.</p> <p>В модуле по большим данным вы узнаете, как строить распределённые пайплайны — это ключевой навык для задач в банке, телекоме и других отраслях, работающих с большими объёмами данных.</p> <p>В модуле по NLP вы освоите классические текстовые представления, а затем переходите к трансформерам и большим языковым моделям, RAG-системам, тонкой настройке моделей и мультимодальным моделям — что позволяет претендовать на роли, связанные с чат-ботами, интеллектуальными поисковыми системами и анализом текстов.</p>
02	Проектный месяц	<p>Командная практика на реальных кейсах. Под руководством опытного наставника выполните и защитите проект, который основан на реальной бизнес-задаче.</p> <p>Примеры задач:</p> <ul style="list-style-type: none">• Предсказание оттока абонентов: анализ телеком-логов и обращений клиентов для удержания пользователей.• Финансовый скоринг как сервис: ML-модель для оценки кредитных рисков с деплоем и мониторингом.• Умная модерация контента: автоматическая фильтрация и анализ текстов пользователей в соцсетях.• Система аналитики для e-commerce: прогнозирование покупок и обработка больших логов пользователей в Spark.
03	Эксперты Школы Анализа Данных от Яндекс	Модуль «Обработка текстов и LLM» разработан совместно с приглашёнными экспертами ШАД — ведущими экспертами IT-индустрии.

ML-инженер

ML-инженер

12 месяцев продолжительность курса
18 проектов в портфолио
18 воркшопов и 5 Q&A-сессий с наставником

ML-инженер расширенный

16 месяцев продолжительность курса
25 проектов в портфолио
21 воркшоп и 5 Q&A-сессий с наставником

Базовый тариф

6 ЧАСОВ

00

Вводный модуль

2-6 НЕДЕЛЬ | 20-70 ЧАСОВ

00

Подготовительный модуль

6 НЕДЕЛЬ | 90 ЧАСОВ

01

Анализ данных с помощью SQL

8 НЕДЕЛЬ | 120 ЧАСОВ

02

Анализ данных с помощью Python

10 НЕДЕЛЬ | 150 ЧАСОВ

03

Классические алгоритмы в машинном обучении

10 НЕДЕЛЬ | 150 ЧАСОВ

04

Расширенные подходы машинного обучения и нейросети

10 НЕДЕЛЬ | 150 ЧАСОВ

05

Автоматизация и внедрение модели в продакшен

2 НЕДЕЛИ | 30 ЧАСОВ

06

Дипломный проект

07

Мастерская
Карьерный трек

Блок недоступен при оплате курса работодателем

Расширенный тариф

2 НЕДЕЛИ | 60 ЧАСОВ

06

Работа с большими данными

8 НЕДЕЛЬ | 120 ЧАСОВ

07

Обработка текстов и LLM

4 НЕДЕЛИ | 60 ЧАСОВ

08

Проектный месяц

2 НЕДЕЛИ | 30 ЧАСОВ

09

Дипломный проект

6 недель | 90 часов
1 итоговый проект

Сформируете навыки извлечения, обработки и анализа данных с помощью SQL. Освоите полный цикл работы с данными: от базовой обработки и очистки до проведения исследовательского анализа и решения бизнес-задач.

Содержание модуля

01. Обработка данных в SQL	<ul style="list-style-type: none">• Введение в машинное обучение• Связи между таблицами• Работа с пропущенными значениями и дубликатами• Присоединение таблиц• Операции множеств и подзапросы• Категоризация значений. Создание новых столбцов• Работа с датой и временем	Практика Написание SQL-запросов в рамках задачи по подготовке данных Встреча с куратором Подготовка к обучению и знакомство со студентами Воркшоп Структурируем сложные SQL-запросы	2 недели
02. Анализ данных и решение ad-hoc задач	<ul style="list-style-type: none">• Знакомство с базами данных• Оконные функции. Агрегирующие функции• Оконные ранжирующие функции• Оконные функции смещения• Описательная статистика. Аналитические функции• Практика решения ad-hoc задач	Проект Решение аналитических запросов различной сложности с помощью SQL в рамках задачи исследовательского анализа данных Воркшоп Применяем алгоритм решения ad-hoc задач на практике	2 недели
03. Итоговый проект модуля	Закрепите навыки обработки данных: очистите и подготовьте их для анализа, соберёте витрину данных для решения задачи бизнеса, оформите отчёт.	Сессия вопросов и ответов с наставником	2 недели

Анализ данных с помощью Python 02

8 недель | 90 часов

1 итоговый проект

Научитесь анализировать и обрабатывать данные с помощью Python: работать с pandas для предобработки данных, проводить исследовательский анализ с использованием ключевых библиотек и визуализировать результаты. Познакомитесь с профессиональными инструментами разработки и основами объектно-ориентированного программирования.

Содержание модуля

04. Предобработка данных в Python	<ul style="list-style-type: none">• Основы библиотеки pandas. Обзор данных• Типы данных. Работа с датой и временем• Индексация в датафреймах. Фильтрация данных• Предподготовка данных• Категоризация данных	Практика Подготовка сырых данных для анализа с помощью Python Воркшоп Обрабатываем пропуски и создаём новые признаки	2 недели
05. Исследовательский анализ данных и визуализация с Python	<ul style="list-style-type: none">• Объединение датафреймов. Срезы данных• Описательная статистика• Взаимосвязь переменных• Визуализация для изучения данных• Сводные таблицы• Пример исследовательского анализа данных	Проект Исследовательский анализ данных для решения бизнес-кейса и подготовка отчёта по исследованию Воркшоп Ищем скрытые закономерности в данных	2 недели
06. Инструменты разработки и основы ООП	<ul style="list-style-type: none">• Работа с терминалом• Виртуальное окружение• Система контроля и управления версиями• Основы ООП	Практика Создание класса предобработки данных, используя принципы ООП. Размещение первого проекта в Git-репозитории Воркшоп Командная работа над ML-проектом через Git	2 недели
07. Итоговый проект модуля	Проведёте исследовательский анализ и визуализируете результаты. Оформите своё исследование в виде структурированного Python-приложения и полноценного репозитория на GitHub	Сессия вопросов и ответов с наставником	2 недели

10 недель | 150 часов

Научитесь переводить бизнес-задачи в ML-формат, готовить данные и выполнять полный цикл построения и оценки модели — от выбора метрик и кросс-валидации до формулировки выводов в бизнес-контексте. Освоите широкий спектр алгоритмов: линейные модели, SVM, KNN и деревья решений — научитесь управлять их сложностью, применять регуляризацию и автоматизированный подбор гиперпараметров для борьбы с переобучением. На практике разберётесь, как усиливать качество решений с помощью ансамблевых методов, использовать современные библиотеки бустинга и логировать эксперименты с MLflow.

Содержание модуля

08. Подготовка данных для машинного обучения	<ul style="list-style-type: none">• Перевод бизнес-задачи в ML-формат• Разделение данных• Кодирование и нормализация признаков• Базовый feature engineering	Проект Перевод бизнес-проблемы в ML-формат, подготовка данных для обучения моделей Воркшоп Проводим комплексный feature engineering	2 недели
09. Первые модели и оценка качества	<ul style="list-style-type: none">• Выбор метрик качества• Обучение линейной регрессии• Обучение логистической регрессии• Проведение кросс-валидации	Проект Обучение базовых моделей и оценка их качества с помощью кросс-валидации Воркшоп Ищем баги: диагностика ML-пайплайна	2 недели
10. Продвинутое линейные модели и регуляризация	<ul style="list-style-type: none">• Переобучение и недообучение• Регуляризация• Метод опорных векторов• Многоклассовая классификация	Проект Применение продвинутых линейных моделей в задачах классификации Воркшоп Боремся с переобучением на практике	2 недели

Содержание модуля

11.	Нелинейные модели и деревья решений	<ul style="list-style-type: none">• k-ближайших соседей (KNN)• Деревья решений• Управление сложностью моделей• Автоматический подбор гиперпараметров с Optuna• Анализ важности признаков	Практика Обучение нелинейных моделей, контроль сложности и оптимизации гиперпараметров Воркшоп Строим деревья решений и анализируем важность признаков	2 недели
12.	Ансамблевые методы	<ul style="list-style-type: none">• Концепция ансамблей и дилемма смещения-разброса• Бэггинг и Random Forest• MLflow для логирования экспериментов• Бустинг и современные библиотеки• Стэкинг моделей	Проект Применение ансамблевых методов и трекинг экспериментов Воркшоп Мастер-класс по MLflow	2 недели

Расширенные подходы машинного обучения и нейросети 04

10 недель | 150 часов
1 итоговый проект

Научитесь применять продвинутые методы машинного обучения и базовые архитектуры нейронных сетей для решения практических задач анализа данных: осознанно выбирать модель в зависимости от типа данных и постановки задачи, обучать и настраивать более сложные алгоритмы, интерпретировать их результаты, понимать ограничения нейросетевых подходов и использовать методы обучения без учителя как дополнительный инструмент работы с данными.

Содержание модуля

13.	Продвинутый отбор признаков и интерпретация предсказаний модели	<ul style="list-style-type: none">• Автоматический отбор признаков• Отбор признаков и сравнение экспериментов в MLflow• Интерпретация моделей с SHAP и сохранение результатов• Работа с дисбалансом классов	Проект Автоматический отбор признаков, интерпретация моделей и управление ML-экспериментами Воркшоп Выбираем устойчивый набор признаков: эксперименты, интерпретация и работа с дисбалансом	2 недели
14.	Обучение без учителя	<ul style="list-style-type: none">• Задачи обучения без учителя• Снижение размерности• Алгоритмы кластеризации• Оценка качества кластеризации	Практика Применение методов обучения без учителя для задачи кластеризации Воркшоп Используем методы снижения размерности на практике	2 недели

Содержание модуля

15.	Введение в нейронные сети и основы PyTorch	<ul style="list-style-type: none">• Введение в нейронные сети• Обучение нейронных сетей• Введение в PyTorch	Проект Построение полного цикла обучения нейросети в PyTorch, включая подготовку данных, реализацию тренировочного и валидационного контуров и работу с гиперпараметрами Воркшоп Ревью первого PyTorch-проекта	2 недели
16.	Рекуррентные нейросети и трансформерные модели	<ul style="list-style-type: none">• Работа с последовательными данными• Простые RNN: теория и практика• Продвинутое RNN: GRU, LSTM, Bidirectional• Основы механизма внимания и трансформеров	Проект Работа с последовательными моделями, обработка текстовых данных и сравнение разных архитектур для задачи автодополнения Воркшоп LSTM или Трансформер: что выбрать для задачи	2 недели
17.	Итоговый проект модуля	Создайте полноценный ML-проект от анализа бизнес-задачи до презентации готового решения. Примените все изученные навыки: извлечение данных из SQL, полный цикл разработки модели, документирование в Git и защита проекта	Сессия вопросов и ответов с наставником	2 недели

Автоматизация и внедрение модели в продакшен

05

10 недель | 150 часов
1 итоговый проект

Научитесь создавать ML-решения, работающие в продакшене: структурировать ML-процессы в воспроизводимые пайплайны с версионированием (DVC) и оркестрацией (Airflow), разворачивать модели через REST API (FastAPI), контейнеризовать сервисы (Docker), проводить A/B-тестирование и внедрять мониторинг моделей в продакшене.

Содержание модуля

18.	Построение ML-пайплайнов и оркестрация	<ul style="list-style-type: none">• Структурированные ML-процессы с sklearn.pipeline• Версионирование и воспроизводимость с DVC• Настройка и оркестрация ML-процессов с Airflow• Интеграция процессов	Проект Создание автоматизированного ML-пайплайна с версионированием и оркестрацией Воркшоп Лаборатория воспроизводимости: DVC vs «просто папки»	2 недели
-----	----------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------

Содержание модуля

19.	Разработка веб-сервисов на базе REST API	<ul style="list-style-type: none">• Введение в веб-разработку для ML-инженеров• REST API: принципы и HTTP-методы• Создание ML-сервисов с FastAPI• Тестирование API	Проект Создание REST API для ML-модели с полным тестированием Воркшоп Разбор готовых FastAPI-сервисов	2 недели
20.	Контейнеризация веб-сервисов	<ul style="list-style-type: none">• Контейнеризация и роль Docker в ML-проектах• Работа с Docker: образы, контейнеры, Dockerfile• Основы автоматизации процессов разработки• Конфигурации автоматизационных пайплайнов	Проект Контейнеризация приложения с Docker Воркшоп Разбираем полный путь контейнера: от кода до деплоя	2 недели
21.	A/B-тестирование и мониторинг модели	<ul style="list-style-type: none">• A/B тестирование в ML: принципы и планирование• Статистический анализ результатов A/B тестов• Системы мониторинга в Prometheus• Визуализация и дашборды для ML-метрик	Проект Проведение A/B тестирования и настройка мониторинга с Prometheus/Grafana Воркшоп Планируем A/B-тесты правильно	2 недели
22.	Итоговый проект модуля	Интегрируете все изученные навыки в комплексном проекте. Создадите полную ML-систему с пайплайном обработки данных, API для модели, контейнеризацией и мониторингом.	Сессия вопросов и ответов с наставником	2 недели

■ Дипломный проект базового тарифа

2 недели | 30 часов

Пройдёте весь жизненный цикл ML-проекта, учитывая потребности бизнеса и применяя полученные навыки

Содержание модуля

Дипломный проект	<ul style="list-style-type: none">• Сбор и предобработка данных• Реализация и обучение модели, исходя из бизнес-требований и ограничений• Подготовка модели к деплою в веб-среде• Упаковка модели в Docker-контейнер• Настройка системы мониторинга производительности модели• Оформление проекта в GitHub-репозиторий	Сессия вопросов и ответов с наставником
------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------

Расширенный тариф

2 недели | 60 часов

Научитесь работать с распределёнными вычислениями в Apache Spark: освоите принципы Big Data и архитектуру распределённых систем, научитесь строить ETL-пайплайны с PySpark (DataFrame API, Spark SQL, UDF) и применять продвинутые трансформации для обработки больших объёмов данных.

Содержание модуля

23. Основы работы с большими данными	<ul style="list-style-type: none">• Big Data: особенности, архитектура и принципы распределённых систем• Экосистема Big Data и инструменты обработки• Форматы данных и основные паттерны хранения• Первые шаги в PySpark	Проект Построение ETL-пайплайна анализа логов веб-сайта в PySpark	2 недели
24. Обработка данных в Apache Spark	<ul style="list-style-type: none">• Архитектура и принципы работы Apache Spark• Spark DataFrame API• Spark SQL• Обработка вложенных структур• UDF и pandas UDF	Проект Построение аналитического пайплайна для анализа поведения пользователей Воркшоп UDF-турнир: когда встроенные функции не спасают	2 недели

Расширенный тариф

8 недель | 120 часов

Научитесь работе с NLP-задачами и современными языковыми моделями: от классических методов представления текста (TF-IDF, Word2Vec, FastText) до глубокого понимания архитектуры трансформеров и LLM, построения RAG-систем с векторным и полнотекстовым поиском, применения методов дообучения (LoRA, QLoRA, SFT, DPO) и работы с мультимодальными моделями (CLIP, SigLIP, VQA).

Содержание модуля

25. Классические методы NLP и подготовка текстов	<ul style="list-style-type: none">• Препроцессинг текста: токенизация, очистка, стоп-слова, лемматизация и стемминг• Статистические представления текста: Bag-of-Words и TF-IDF• Семантические эмбединги: Word2Vec и FastText	Проект Классический NLP-пайплайн для классификации отзывов: от очистки текста и токенизации до векторизации, использования моделей и аналитического сравнения подходов	2 недели
26. Архитектура трансформеров и устройство LLM	<ul style="list-style-type: none">• Внутреннее устройство трансформера• От encoder/decoder к LLM• Производительность и инференс	Проект Практическое сравнение трёх архитектур трансформеров Воркшоп Encoder vs Decoder: выбираем модель под задачу	2 недели
27. Построение RAG-систем и методы поиска	<ul style="list-style-type: none">• Введение в RAG• Векторный поиск и векторные базы данных• Retrieval методы• Полнотекстовый поиск и дедупликация	Проект Создание RAG-системы с векторным поиском и ранжированием	2 недели
28. Мультимодальность и дообучение LLM	<ul style="list-style-type: none">• Методы адаптации LLM: pretrain, SFT и DPO• Дообучение LLM• Мультимодальные модели• Анализ качества мультимодальных данных	Проект Построение мультимодальной системы поиска: подготовить эмбединги изображений, реализовать функцию поиска по тексту и продемонстрировать релевантные результаты Воркшоп Дообучение LLM на ограниченных ресурсах: LoRA vs QLoRA	2 недели

Расширенный тариф

4 недели | 60 часов

Под руководством опытного наставника выполните и защитите проект, который основан на реальной бизнес-задаче

Дипломный проект расширенного тарифа

09

Расширенный тариф

2 недели | 30 часов

Пройдёте весь жизненный цикл ML-проекта, учитывая потребности бизнеса и применяя полученные навыки

Содержание модуля

Дипломный проект

- Сбор и предобработка данных
- Реализация и обучение модели, исходя из бизнес-требований и ограничений
- Подготовка модели к деплою в веб-среде
- Упаковка модели в Docker-контейнер
- Настройка системы мониторинга производительности модели
- Оформление проекта в GitHub-репозиторий

Сессия вопросов и ответов
с наставником