

# Специалист по Data Science

Продолжительность курса — 8 месяцев, 380 часов

15 проектов в портфолио

00

## Вводная часть

Основы Python  
и анализа данных

От 8 часов

01

## Введение в анализ данных

Базовый Python  
Предобработка данных  
Исследовательский анализ  
данных  
Статистический анализ данных  
Сборный проект

10 недель, включая каникулы  
(1 неделя)

5 проектов в портфолио

02

## Основы машинного обучения

Линейные модели  
в машинном обучении  
Обучение с учителем:  
качество модели  
Сборный проект

8 недель, включая каникулы  
(1 неделя)

3 проекта в портфолио

03

## Продвинутое машинное обучение

Машинное обучение в бизнесе  
Базовый SQL  
Численные методы  
Временные ряды

8 недель, включая каникулы  
(1 неделя)

4 проекта в портфолио

04

## Машинное обучение для больших объёмов данных

Машинное обучение для текстов  
Компьютерное зрение  
Обучение без учителя

6 недель, включая каникулы  
(1 неделя)

2 проекта в портфолио

## Выпускной проект

2 недели

1 проект в портфолио

## Дополнительные (необязательные) спринты

Практика Python  
Теория вероятностей  
Практика SQL

# 00

## Бесплатная вводная часть

от 8 часов

### Основы Python и анализа данных

Узнаете основные концепции анализа данных и поймёте, чем занимаются аналитики данных и специалисты по Data Science.

#### Решите задачи из разных областей:

- выясните причину массовой поломки гаджетов;
- проверите, как новая опция в видеоигре влияет на пользователей;
- поможете кофейной компании классифицировать клиентов;
- оцените эффективность роботов в службе поддержки.

Решая кейсы, вы изучите азы Python и библиотеки pandas, научитесь строить некоторые графики и верно их трактовать.

#### Темы

- Moscow Catnamys
- Ошибки в коде
- Переменные и типы данных
- Как выдвигать гипотезы
- Что делают специалисты в области данных
- Машинное обучение
- Финальный проект



# 01

## Введение в анализ данных

2 недели  
от 30 часов

### Базовый Python

Глубже погрузитесь в язык программирования Python и работу с библиотекой pandas.

#### Темы

- Переменные и типы данных
- Строки
- Списки
- Цикл for
- Вложенные списки
- Условный оператор. Цикл while
- Функции
- Словари
- Библиотека pandas
- Предобработка данных
- Анализ данных и оформление результатов
- Jupyter Notebook — тетрадь в ячейку

#### Проект

Сравните данные пользователей Яндекс Музыки по городам и дням недели.

2 недели  
от 35 часов

### Предобработка данных

Научитесь очищать данные от выбросов, пропусков и дубликатов, а также преобразовывать разные форматы данных.

#### Темы

- Введение в предобработку данных
- Работа с пропусками
- Изменение типов данных
- Поиск дубликатов
- Категоризация данных
- Системное и критическое мышление в работе аналитика

#### Проект

Проанализируете данные о клиентах банка и определите, какие показатели влияют на их кредитоспособность.



2 недели  
от 30 часов

## Исследовательский анализ данных

Изучите основы теории вероятностей и статистики. Примените их для исследования свойств данных, поиска закономерностей, распределений и аномалий. Познакомьтесь с библиотеками SciPy и Matplotlib. Создадите диаграммы, поупражняйтесь в анализе графиков.

### Темы

- Введение в исследовательский анализ данных
- Первые графики и выводы
- Изучение срезов данных
- Работа с несколькими источниками данных
- Взаимосвязь данных
- Валидация результатов

### Проект

Исследуете архив объявлений о продаже объектов недвижимости в Санкт-Петербурге и Ленинградской области.

2 недели  
от 20 часов

## Статистический анализ данных

Научитесь анализировать взаимосвязи в данных методами статистики. Узнаете, что такое статистическая значимость, гипотезы и доверительные интервалы.

### Темы

- Комбинаторика
- Теория вероятностей
- Описательная статистика
- Случайные величины
- Распределения
- Проверка гипотез

### Проект

Проверите гипотезы сервиса аренды самокатов, чтобы помочь вырастить бизнес.

1 неделя  
от 15 часов

## Сборный проект

Научитесь предварительному исследованию данных, сформулируете и проверите гипотезы.

### Проект

Найдёте закономерности в данных о продаже игры.

Каникулы —————> 1 неделя

# 02

## Основы машинного обучения

3 недели  
от 35 часов

### Линейные модели в машинном обучении

Познакомьтесь с библиотекой scikit-learn и циклом создания модели машинного обучения. Освойте линейные модели. Научитесь решать задачи регрессии и классификации. Изучите основы линейной алгебры. Узнаете, как подготовить данные для машинного обучения и проверить работу модели.

#### Темы

- Моделирование и машинное обучение
- Первая модель и библиотека scikit-learn
- Линейная алгебра и библиотека numpy
- Метрики в задаче регрессии
- Скрытые ошибки модели
- От регрессии к классификации
- Метрики в задаче классификации
- Схема моделирования

#### Проект

Разработаете модель машинного обучения, которая поможет владельцам молочной фермы принимать эффективные решения.

3 недели  
от 35 часов

### Обучение с учителем: качество модели

Изучите новые модели — SVM и kNN. Научитесь настраивать модели машинного обучения. Освойте новые метрики качества решений модели и подходы для их улучшения. Узнаете, что такое пайплайны и как они упрощают разработку моделей.

#### Темы

- Метод опорных векторов
- Переобучение и регуляризация
- kNN
- Дисбаланс классов
- Дерево решений и гиперпараметры
- Работа с признаками
- Отбор признаков и интерпретация моделей
- Пайплайн обучения

#### Проект

Проанализируете покупки клиентов в ритейл-сети. Разработаете модель, которая поможет маркетологам запускать эффективные акции.



1 неделя  
от 20 часов

## Сборный проект

Разработаете две модели машинного обучения и оцените их качество. Упакуете работу в пайплайн.

Проект

Смоделируете коэффициент удовлетворённости сотрудников, чтобы помочь HR-отделу компании спрогнозировать текучку кадров.

Каникулы —————> 1 неделя



# 03

## Продвинутое машинное обучение

2 недели  
от 25 часов

### Машинное обучение в бизнесе

Узнаете, как машинное обучение (МО) помогает бизнесу, как собирать данные и как продуктовые метрики связаны с показателями МО. Научитесь запускать новую функциональность сервиса, применяя МО. Узнаете, что такое бизнес-метрики, KPI и A/B-тестирование.

#### Темы

- Введение в машинное обучение в бизнесе
- Метрики бизнеса
- Запуск новой функциональности
- Сбор данных
- Поведенческие алгоритмы

#### Проект

Обучите модель и предскажете, в каком месторождении добыча нефти принесёт наибольшую прибыль.

3 недели  
от 30 часов

### Базовый SQL

Изучите основы структурированного языка запросов SQL и операции реляционной алгебры. Познакомьтесь с PostgreSQL — популярной системой управления базами данных (СУБД). Научитесь писать запросы разного уровня сложности и переводить бизнес-задачи на язык SQL. Также вы познакомитесь с PySpark — библиотекой с открытым исходным кодом, которая применяется для распределённой обработки больших объёмов данных.

#### Темы

- Введение в базы данных
- Срезы данных в SQL
- Агрегирующие функции. Группировка и сортировка данных
- Взаимоотношения между таблицами. Типы объединений таблиц
- Подзапросы и временные таблицы
- PySpark
- Схемы данных и оконные функции

#### Проект

Напишите ряд запросов разной сложности к базе данных, в которой хранится информация о венчурных инвесторах, стартапах и инвестициям в них.



1 неделя  
от 10 часов

## Численные методы

Разберёте ряд алгоритмов и приспособите их к решению практических задач с использованием численных методов. Освоите приближённые вычисления, оценки сложности алгоритма, градиентный спуск. Узнаете, как обучаются нейронные сети и что такое градиентный бустинг.

### Темы

- Анализ алгоритмов
- Градиентный спуск
- Обучение градиентным спуском
- Градиентный бустинг
- Паттерны решения задачи

### Проект

Разработаете модель для определения стоимости автомобиля с пробегом.

1 неделя  
от 10 часов

## Временные ряды

Временные ряды описывают, как параметры, например объём потребления электроэнергии или количество заказов такси, меняются с течением времени. Вы научитесь анализировать ряды, искать тренды и выявлять сезонность. Узнаете, как создавать табличные данные и задачу регрессии из временных рядов.

### Темы

- Введение во временные ряды
- Анализ временных рядов
- Прогнозирование временных рядов

### Проект

Построите модель и спрогнозируете пиковые нагрузки в такси.

Каникулы —————> 1 неделя





# 04

## Машинное обучение для больших объёмов данных

2 недели  
от 25 часов

### Машинное обучение для текстов

Научитесь преобразовывать тексты и решать для них задачи классификации и регрессии. Узнаете, как с помощью различных подходов вычислять эмбединги — векторные представления слов.

#### Темы

- Введение в машинное обучение для текстов
- Векторизация текстов
- Языковые представления

#### Проект

Ускорите модерацию комментариев в сообществе, автоматизировав оценку их токсичности.

2 недели  
от 25 часов

### Компьютерное зрение

Научитесь решать простые задачи компьютерного зрения с привлечением готовых нейронных сетей и библиотеки Keras. Заглянете в мир глубокого обучения.

#### Темы

- Введение в компьютерное зрение
- Полносвязные сети
- Свёрточные сети

#### Проект

Постройте модель для определения приблизительного возраста человека по фотографии.

1 неделя  
от 5 часов

### Обучение без учителя

Обучение без учителя — это один из способов машинного обучения, при котором система решает задачу без размеченных заранее данных, на основе их особенностей и структуры. Познакомитесь с задачами кластеризации и поиска аномалий.

#### Темы

- Введение в обучение без учителя
- Кластеризация
- Поиск аномалий

Каникулы —————> 1 неделя

# Выпускной проект

2 недели  
от 30 часов

В последнем проекте подтвердите, что освоили новую профессию. Уточните задачу заказчика, пройдёте все стадии анализа данных и машинного обучения. Теперь без уроков и домашних заданий — всё как на реальной работе.

Выполните один из проектов:

## Проект

1. Спрогнозируете отток клиентов в телекоммуникационной компании.
2. Предскажете параметры технологического процесса на металлургическом комбинате.

## Дополнительные курсы

Это необязательные спринты. Каждый студент сам выбирает один из вариантов прохождения:

— Освоить дополнительный курс из нескольких коротких уроков, освежить в памяти теорию и решить задачи.

— Пропустить курс совсем или вернуться к нему, когда будут время и необходимость.

### Практика Python

Дополнительные задачи для тренировки индексов, циклов, логических значений, условного оператора и фильтрации на Python.

### Теория вероятностей

Вспомните или узнаете базовые термины теории вероятностей: независимые, противоположные, несовместные события и т. д. На простых примерах и забавных задачах потренируетесь работать с числами и выстраивать логику решения. После попробуете себя в решении задач для собеседований.

### Практика SQL

Несколько десятков дополнительных задач на отработку навыка работы с SQL.



# Карьерный трек

В дополнение к основной образовательной программе вы можете пройти трек подготовки к трудоустройству, чтобы быстрее найти новую работу. Программа включает руководства, вебинары и домашние задания.

## 1. Резюме

- Изучите подходы к составлению резюме.
- Научитесь презентовать свой опыт работы.
- Составите резюме вместе со специалистом по трудоустройству.

## 2. Портфолио и сопроводительное ПИСЬМО

- Научитесь создавать портфолио на сервисе GitHub.
- Изучите особенности написания сопроводительного письма.
- Напишите сопроводительное письмо вместе со специалистом по трудоустройству.

## 3. Интервью

Научитесь презентовать себя и легче проходить собеседования.

## 4. Поиск работы

- Узнаете, где искать работу.
- Разберёте активную и пассивную стратегии поиска работы.
- Научитесь бороться со страхами во время трудоустройства.

## 5. Активный поиск работы

- Начнёте погружаться в профессиональное сообщество.
- Получите доступ к партнёрским вакансиям.
- Поучаствуете в проектах от реальных заказчиков.