

# Кто такие специалисты по Data Science

Продолжительность курса — 5 месяцев

## 01. В чём суть профессии

Специалисты по Data Science помогают компаниям принимать решения на основе данных. Они решают технические задачи: создают модели для прогнозов и пользуются более сложными инструментами, основанными на математике.

## 02. В каких сферах работают

Data Science востребована в маркетинге, финансах, промышленности, разработке любых новых продуктов и технологий.

## 03. Какие задачи решают

Банк хочет автоматически прогнозировать платёжеспособность новых клиентов. Нужно использовать данные о текущих клиентах банка.

Поисковому сервису нужен алгоритм, который поможет определять фотографии с запрещённым содержанием и исключать их из поисковой выдачи.

Онлайн-кинотеатр построит рекомендательную систему фильмов и сериалов для пользователей.

## Зарплатная вилка



## На курсе — всё, что нужно, чтобы начать карьеру специалиста по DS

Диплом о переподготовке

Непрерывная практика

Портфолио из 15 проектов

Помощь в трудоустройстве

# Почему этот курс подойдёт студентам без опыта

## Свободное расписание

Читать теорию и практиковаться можно в любое время, главное — соблюдать сроки сдачи проектов. У нас есть мобильное приложение для учёбы, чтобы вы не были привязаны к компьютеру.

## Понятная теория

Термины и правила подкреплены примерами из жизни. Сложность и длина программы курса рассчитаны так, чтобы каждую следующую главу вы понимали всё лучше. Мы постоянно обновляем программу: последнее обновление на данный момент — апрель 2023 года

## Практика в тренажёре

Учитесь читать, визуализировать и интерпретировать данные. Ошибайтесь, быстро получайте обратную связь и исправляйте ошибки

## Учёба на реальных задачах

Вас ждут типичные для аналитика задачи из разных сфер. Вы решите их и сможете добавить в своё портфолио.

## Команда сопровождения

Мы поддержим, объясним сложные темы, поможем улучшить проекты и не дадим сдать на полпути.

Python и его библиотеки

SQL — универсальный язык  
запросов к базам данных

Математика и статистика

Jupyter Notebook для экспериментов и  
разведочного анализа данных

Scikit-learn, CatBoost для создания  
моделей и машинного обучения

## Чтобы начать учиться, не нужен опыт в IT

Среди наших выпускников, которые стали специалистами по DS, есть спортсмены, экономисты, таксисты и другие специалисты. После наших курсов аналитики находят работу и в стартапах, и в корпорациях — например, в Яндексе, «Ренессанс Банке» и Wildberries.

# Data Science Bootcamp



## Вводный курс: основы Python и анализа данных

Узнаете основные концепции анализа данных и поймёте, чем занимаются аналитики данных и специалисты по Data Science. Решите пять кейсов по работе с данными из разных областей: **от 20 часов**

- выясните причину массовой поломки гаджетов;
- проверите окупаемость рекламы мобильного приложения;
- поможете выбрать стратегию развития ИИ-стартапа;
- оцените эффективность роботов в службе поддержки.

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

## Базовый Python

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

1 неделя, от 30 часов

### 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: музыка больших городов

### Проект

Сравнение данных пользователей Яндекс.Музыки по городам и дням недели.

# Предобработка данных

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

1 неделя, от 35 часов

## 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.

# Исследовательский анализ данных

Изучите основы теории вероятностей и статистики. Примените их для исследования основных свойств данных, поиска закономерностей, распределений и аномалий. Познакомитесь с библиотеками SciPy и Matplotlib. Отрисуете диаграммы, поупражняетесь в анализе графиков.

1 неделя, от 30 часов

## 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.

# Статистический анализ данных

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

1 неделя, от 20 часов

## Темы

- Введение в статистический анализ данных
- Описательная статистика
- Теория вероятностей
- Проверка гипотез

## Проект

Оптимизируете воронки продаж, чтобы ускорить работу отдела маркетинга.

# Итоговый проект первого модуля

Научитесь предварительному исследованию данных, сформулируете и проверите гипотезы.

1 неделя, от 15 часов

## Проект

Найдёте закономерности в данных о продаже игры.

# Введение в машинное обучение

Освойте основные концепции машинного обучения. Познакомьтесь с библиотекой Scikit-Learn и примените её для создания первого проекта с машинным обучением.

1 неделя, от 20 часов

## Темы

- Введение в курс
- Первая обученная модель
- Качество модели
- Улучшение модели
- Переходим к регрессии

## Проект

Разработаете систему рекомендации тарифов для оператора мобильной связи.

# Обучение с учителем: классификация и регрессия

Углубитесь в самую востребованную область машинного обучения – обучение с учителем. Узнаете, как обращаться с несбалансированными данными.

1 неделя, от 20 часов

## Темы

- Введение в обучение с учителем
- Подготовка признаков
- Метрики классификации
- Несбалансированная классификация
- Метрики регрессии
- Поведенческие алгоритмы

## Проект

Спрогнозируете вероятность ухода клиента из банка.

# Машинное обучение в бизнесе

Узнаете, как машинное обучение (МО) помогает бизнесу, как собирать данные и как продуктовые метрики связаны с показателями МО.

1 неделя, от 25 часов

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.

# Итоговый проект второго модуля

Подготовьте данные для машинного обучения. С помощью модели оцените её качество.

1 неделя, от 20 часов

## Проект

Подготовьте данные для машинного обучения. С помощью модели оцените её качество.

# Линейная алгебра

Заглянете внутрь нескольких изученных ранее алгоритмов и лучше поймёте, как их применять. На практике освоите с нуля главные концепции линейной алгебры: линейные пространства, линейные операторы, евклидовы пространства.

1 неделя, от 20 часов

## Темы

- Введение в линейную алгебру
- Векторы и векторные операции
- Расстояние между векторами
- Матрицы и матричные операции
- Линейная регрессия изнутри

## Проект

Методом преобразования данных защитите личную информацию клиентов страховой компании.

# 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.

1 неделя, от 20 часов

## Темы

- Введение в численные методы
- Анализ алгоритмов
- Градиентный спуск
- Обучение градиентным спуском
- Градиентный бустинг
- Поведенческие алгоритмы

## Проект

Разработаете модель для определения стоимости автомобиля с пробегом.

# Временные ряды

Научитесь анализировать временные ряды. Узнаете, как создавать табличные данные из временных рядов и решать для них задачу регрессии.

1 неделя, от 15 часов

## Темы

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

## Проект

Построите модель и спрогнозируете пиковые нагрузки в такси.

# Машинное обучение для текстов

Научитесь делать числовые векторы из текстов и решать для них задачи классификации и регрессии. Узнаете, как вычисляются признаки TF-IDF и познакомитесь с языковыми представлениями word2vec и BERT.

1 неделя, от 25 часов

## Темы

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

## Проект

Ускорите модерацию комментариев в сообществе, автоматизировав оценку их токсичности.

# 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

1 неделя, от 30 часов

## Проект

Напишите ряд запросов разной сложности к базе данных, в которой хранятся данные по венчурным инвесторам, стартапам и инвестициям в них.

# Компьютерное зрение

Научитесь решать простые задачи компьютерного зрения с привлечением готовых нейронных сетей и библиотеки Keras. Одним глазком заглянете в Deep learning.

## Темы

- Введение в компьютерное зрение
- Полносвязные сети
- Свёрточные сети
- Поведенческие алгоритмы

1 неделя, от 25 часов

## Проект

Постройте модель для определения приблизительного возраста человека по фотографии.

# Выпускной проект

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.

1 неделя, от 30 часов

## Один из проектов:

- Постройте модель, прогнозирующую отток клиентов в телекоммуникационной компании.
- Постройте модель, предсказывающую параметры технологического процесса на металлургическом комбинате.

## Дополнительные курсы

Это необязательные спринты. Значит, 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 будет время и необходимость.

## Теория вероятностей

Вспомните или узнаете базовые термины в 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.



# Трек программы трудоустройства

В дополнение к основной образовательной программе вы можете пройти трек трудоустройства, чтобы быстрее найти новую работу. Программа включает руководства, вебинары, домашние задания. Каждая секция трека длится 2 недели и занимает около 10 часов.

## 1 Резюме

- Изучите подходы к составлению резюме.
- Разберётесь в стратегиях поиска работы.
- Сформируете резюме совместно со специалистом по трудоустройству.

## 2 Портфолио и сопроводительное письмо

- Научитесь создавать портфолио на сервисе GitHub.
- Изучите подходы к написанию сопроводительного письма.
- Сформируете сопроводительное письмо со специалистом по трудоустройству.

## 3 Интервью

- Научитесь презентовать себя и легче проходить собеседования.

## 4 Нетворкинг и поиск работы

- Начнёте погружаться в профессиональное сообщество.
- Узнаете, где искать работу.