

Skillbox

# Профессия Data Scientist с 0 до PRO

# Ключевые особенности программы

- **Две специальности**

В одной программе

- **Реальные данные**

В практических заданиях

- **300 часов**

Онлайн-обучения

- **Доступ к курсам**

Навсегда

- **Трудоустройство**

После обучения

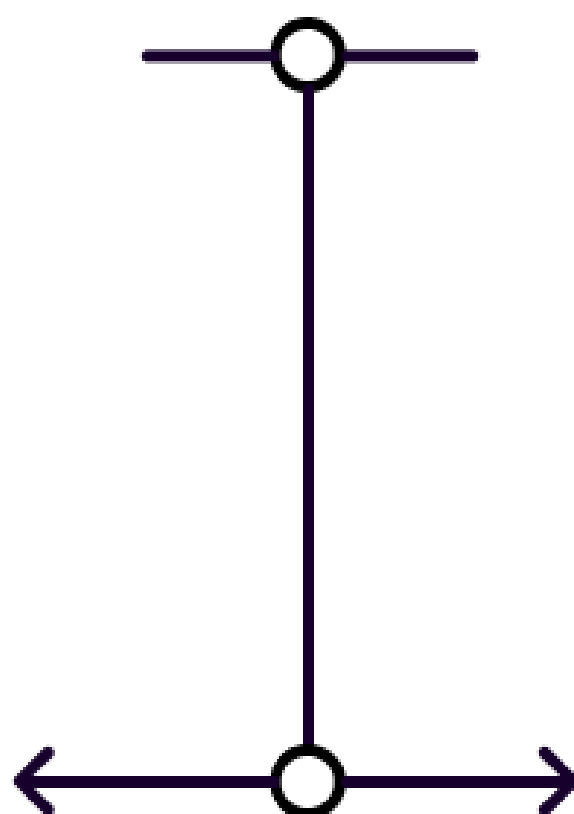
# Структура курса

## УРОВЕНЬ 1

### Введение в анализ данных

## УРОВЕНЬ 1

### Введение в машинное обучение



Специализация на выбор

## УРОВЕНЬ 2

### Специализация: Анализ данных.

- язык программирования R
- средства создания интерактивных дашбордов
- анализом временных рядов

## УРОВЕНЬ 2

### Специализация: Машинное обучение

- нейронные сети
- машинное зрение
- разработка рекомендательных систем
- обработка языка.

Skillbox

УРОВЕНЬ 1

# Введение в анализ данных и машинное обучение

# Введение в анализ данных и машинное обучение

Вы научитесь использовать язык программирования Python, работать с источниками данных: CSV, XML и XLSX, освоите базовые приемы визуализации с помощью Matplotlib и поработаете с разными моделями машинного обучения.

Skillbox

УРОВЕНЬ 1

# **Блок 1. Аналитика. Начальный уровень**

## МОДУЛЬ 1

## Введение в Data Science

- про Data Science;
- авторы курса;
- структура курса;
- практика.

## МОДУЛЬ 2

## Введение в Python

- Причины выбора Anaconda для Python
- Установка окружения
- Jupyter Notebook: интерфейс, ячейки и разметка текста, формулы, изображения, работа в меню
- Python: базовые вычисления, операции с числами
- Переменные в Python: создание и изменение, типы значений и переменных, преобразование типов
- Домашняя работа

## МОДУЛЬ 3

**Библиотека NumPy (ч.1)**

- Основные идеи NumPy
- Nddarray: базовая концепция и операции
- Базовые операции над массивами
- Вычисления с массивами
- Создание массивов
- Условия и булевы массивы
- Домашняя работа

## МОДУЛИ 4

**Модуль 4 Библиотека NumPy (ч.2)**

- Изменение размерности списков
- Изменение списков
- Создание новых списков на основе старых
- Дополнительные концепции NumPy
- Структурированные списки
- Чтение/запись списков
- Домашняя работа



## МОДУЛИ 5

**Библиотека pandas (ч.1)**

- Основные структуры
- Индексы и метод .iloc
- Индексы и метод .loc
- Характеристики датафрейма pandas
- Операции над датафреймами
- Домашняя работа

## МОДУЛЬ 6

**Библиотека pandas (ч.2)**

- Применение функций и метод .apply
- Группировка и агрегирование
- Сортировка и упорядочивание
- Работа с NaN-ами
- Иерархическое индексирование
- Домашняя работа

## МОДУЛИ 7

## Чтение и запись данных (ч.1)

- Чтение текстовых и CSV файлов
- Работа с данными формата HTML
- Работа с данными формата XML
- Работа с данными формата XLSX
- Домашняя работа

## МОДУЛЬ 8

## Основы SQL

- Введение в SQL
- Структура БД, запрос DESCRIBE
- Выбор и фильтрация данных, запрос SELECT
- Функции и выражения, агрегация данных
- Отношения и соединение таблиц: JOIN
- Группировка
- Модификация данных
- Подзапросы
- Объявление и изменение структуры данных
- Домашняя работа

## МОДУЛЬ 9

**Чтение и запись данных (ч.2)**

- JSON
- Pickle - сериализация объектов
- Работа с данными формата HDF5
- Работа с базой данных SQLite3
- Работа с базой данных PostgreSQL
- Работа с базой данных из pandas
- Работа с базой данных MongoDB
- Домашняя работа

## МОДУЛЬ 10

**Работа со строками**

- Стандартные приёмы работы с текстом
- Регулярные выражения python
- Статистики текста
- Домашняя работа

## МОДУЛЬ 11

## Визуализация данных с помощью matplotlib

- Архитектура matplotlib
- Простой график Matplotlib
- Сохранение и дальнейшая работа с графиком
- Добавление элементов на график
- Простые графики Matplotlib
- Сложные графики Matplotlib
- Домашняя работа

## МОДУЛЬ 12

## Заключение

Skillbox

УРОВЕНЬ 1

# Блок 2. Машинное обучение. Начальный уровень

## МОДУЛЬ 1

**Основные концепции Machine Learning (ML)**

- Что такое машинное обучение?
- Обучение с учителем. Регрессия (Supervised Learning)
- Обучение с учителем. Классификация (Supervised Learning)
- Обучение без учителя. Кластеризация (Unsupervised Learning)
- Обучение с подкреплением. Нейронные сети и ансамбли (Reinforcement Learning)
- Домашняя работа

## МОДУЛИ 2

**Жизненный цикл ML-проекта**

- Понимание целей бизнеса, работа с данными.
- Моделирование, развертывание.

## МОДУЛЬ 3

**Основные концепции Machine Learning (ML)**

- Что такое машинное обучение?
- Обучение с учителем. Регрессия (Supervised Learning)
- Обучение с учителем. Классификация (Supervised Learning)
- Обучение без учителя. Кластеризация (Unsupervised Learning)
- Обучение с подкреплением. Нейронные сети и ансамбли (Reinforcement Learning)
- Домашняя работа

## МОДУЛИ 4

**Регрессия (ч.1)**

- Линейная регрессия
- Метрики качества регрессии
- Полиномиальная регрессия
- Домашняя работа

## МОДУЛЬ 5

## Регрессия (ч.2)

- Регуляризация
- Обучение моделей с помощью градиентного спуска
- Домашняя работа

## МОДУЛИ 6

## Классификация (ч.1)

- Алгоритм KNN
- Алгоритм SVM
- Алгоритм Decision Trees
- Домашняя работа

## МОДУЛИ 7

## Классификация (ч.2)

- Метрики качества классификации
- Мультиклассовая классификация
- Анализ ошибок
- Домашняя работа



## МОДУЛЬ 8

## Кластеризация

- Алгоритм DBSCAN
- Алгоритм K-средних
- Метрики качества кластеризации
- Домашняя работа

## МОДУЛИ 9

## Классификация (ч.1)

- PCA
- SVD
- t-SNE
- Домашняя работа

## МОДУЛЬ 10

## Дополнительные техники (ч.2)

- Бустинг моделей
- Стэкинг моделей
- Домашняя работа

## МОДУЛИ 11

## Знакомство с Kaggle

- Kaggle
- Домашняя работа

## МОДУЛИ 12

## Заключение

Skillbox

УРОВЕНЬ 2

# Специализация: аналитика или машинное обучение

# Специализация: аналитика или машинное обучение

Анализ данных. Познакомьтесь с языком R, средствами создания интерактивных дашбордов и анализом временных рядов, что поможет вам развиваться в области BI.

Машинное обучение. Специализация даст вам знания в области нейронных сетей, а также осветит дополнительные задачи Machine Learning: машинное зрение, разработку рекомендательных систем и обработку языка.

Skillbox

УРОВЕНЬ 2

# **Блок 3. Вариант 1. Аналитика.**

## **Средний уровень**

## МОДУЛЬ 1

**Введение**

## МОДУЛЬ 2

**А/В-тестирование**

- АБ-тесты для проверки гипотез
- Оценка параметров по выборке: предсказательный интервал
- Доверительный интервал: точечная оценка параметров по выборке
- Определение объема репрезентативной выборки для АБ-теста
- Механизм статистических гипотез: основные определения
- Практические примеры аналитики АБ-теста
- Домашняя работа

## МОДУЛИ 3

**Анализ временных рядов**

- Способы обработки временных рядов
- Эконометрический подход
- Примеры моделей машинного обучения для предсказания значений временного ряда
- 
- Домашняя работа

## МОДУЛЬ 4

## Выявление аномалий

- Поиск аномалий с помощью Scikit-Learn
- Домашняя работа

## МОДУЛЬ 5

## Язык программирования R (ч.1)

- Установка R
- Работа с консолью
- Типы данных
- Работа с коллекциями: индексирование
- Домашняя работа

## МОДУЛЬ 6

## Язык программирования R (ч.2)

- Управляющие структуры
- Функции
- Область видимости
- Домашняя работа

## МОДУЛЬ 7

## Язык программирования R (ч.3)

- Работа с коллекциями: функции
- Отладка и профилирование
- Простейшая графика
- Домашняя работа

## МОДУЛЬ 8

## Создание аналитических панелей (Dashboard) - R View

- Введение во фреймворк Shiny
- Обзор возможностей ч.1
- Обзор возможностей ч.2
- Обзор возможностей ч.3
- Домашняя работа



## МОДУЛЬ 9

## Создание аналитических панелей (Dashboard) - Python View

- Введение во фреймворк Dash
- Обзор возможностей ч.1
- Обзор возможностей ч.2
- Обзор возможностей ч.3
- Домашняя работа

## МОДУЛЬ 10

## Аналитические СУБД

- База данных Vertica
- База данных Exasol
- База данных Clickhouse

## МОДУЛЬ 11

## Заключение

Skillbox

УРОВЕНЬ 2

**Блок 3. Вариант 2.**  
**Машинное обучение.**  
**Средний уровень**

## МОДУЛЬ 1

**Введение**

## МОДУЛЬ 2

**Введение в нейронные сети (ч.1)**

- Классификация. Логистическая регрессия
- Нейронные сети, функции активации
- Градиентный спуск. Лосс функции
- Домашнее задание

## МОДУЛИ 3

**Введение в нейронные сети (ч.2)**

- Фреймворки для нейронных сетей.
- Автоград, вычислительный граф.
- Особенности TF. Тьюриал по TF и Colab.
- Домашнее задание

## МОДУЛЬ 4

## Сверточные нейросети (классификация)

- Свертки, фильтры. Примеры и интуиция
- Стандартные архитектуры и слои.
- Transfer learning, визуализация фичей

## МОДУЛИ 5

## Семантическая сегментация и детектирование объектов

- Слабая локализация, полносверточные нейросети (FCN)  
Семантическая сегментация: U-net, skipped connections, Transposed conv, Atrous conv
- Демо: построение и обучение нейросети для семантической сегментации
- Локализация, регрессия bounding box'a
- Детектирование объектов: RCNN методы, YOLO/SSD
- Демо: TensorFlow Object Detection API
- Домашнее задание

## МОДУЛЬ 6

**GAN, VAE, Style Transfer**

- Задача генерации текстур. Style transfer.
- Feed forward style transfer.
- GAN, VAE, CycleGan

## МОДУЛИ 7

**NLP на нейросетях: RNN, LSTM, GRU, Attention, Transformer**

- Рекуррентные нейронные сети
- Продвинутые RNN: LSTM/GRU
- Типы обработки последовательностей: one-to-many, many-to-one, many-to-many
- Seq2Seq, Encoder-Decoder
- Механизм внимания (Attention), архитектура Transformer
- Демо: Обучение RNN для классификации текстов
- Демо: Обучение RNN для генерации текстов
- Домашнее задание

## МОДУЛЬ 8

**RL: Обучение с подкреплением**

- Общие теоретические основы обучения с подкреплением
- Q-функция, уравнение Беллмана
- Q-Learning: табличный, аппроксимация, DQN
- Трюки: Exploration, Experience replay
- Другие типы RL
- Демо: Обучение табличной Q-функции
- Демо: Обучение DQN
- Домашнее задание

## МОДУЛИ 9

**Внедрение в Production**

- Ускорение нейронных сетей (обучение и инференс)
- Библиотека TensorRT
- Нейросеть как сервис: TRT inference server, TF serving
- Демо: конвертация TensorFlow модели в TensorRT для быстрого инференса
- Домашнее задание

## МОДУЛЬ 10

## Введение в NLP

- tf-idf
- n-grams
- Модель Bag of Words
- Техника Word2Vec
- Домашняя работа

## МОДУЛИ 11

## Рекомендательные системы

- Рекомендательные системы: цели и задачи
- KNN рекомендации
- Эмбединги пользователей и контента (скрытые факторы)
- Метрики рекомендательных систем
- Домашняя работа

## МОДУЛИ 12

## Вывод моделей машинного обучения в production, post-production и мониторинг

## МОДУЛИ 13

## Заключение



Skillbox

**Преподаватели**



## Валентин Пановский

Chief Data Scientist в Skillbox

### О преподавателе

- кандидат физико-математических наук;
- преподаватель МАИ, факультет «Информационные технологии и прикладная математика»;
- автор и соавтор более 50 научных работ в российских и международных изданиях и двух зарубежных монографий;
- практик в различных областях: от рекомендательных систем в e-commerce до Deep Learning для улучшения качества метеорологических моделей.



## Михаил Овчинников

Руководитель команды разработки внутренних инструментов в Facebook

### О преподавателе

- главный методист технического направления Skillbox;
- экс-руководитель команды разработки систем машинного обучения и антиспама в Badoo;
- занимается разработкой ПО последние 15 лет;
- докладчик крупнейших IT-конференций России;
- специалист в области разработки высоконагруженных систем, обработки больших данных и машинного обучения.



## Алла Тамбовцева

Преподаватель НИУ ВШЭ

### О преподавателе

- преподаватель программы «Python для анализа данных» Центра непрерывного образования факультета компьютерных наук НИУ ВШЭ;
- разработчик курсов «Основы программирования в R», «Программирование для анализа данных», Data Analysis in the Social Sciences.



## Александр Джумурат

Руководитель команды разработки рекомендательной системы в ivi.ru

### О преподавателе

- разработчик курсов «Управление данными» и «Анализ данных» магистерской программы МАИ;
- разработчик адаптивных алгоритмов обучения в LMS class.net МАИ;
- автор корпоративного блога ivi на habr.com.



## Дмитрий Коробченко

Deep Learning R&D инженер и менеджер в NVIDIA;

### О преподавателе

- работал инженером в IBM и руководителем проектов в Samsung Electronics;
- эксперт в области нейронных сетей, машинного обучения и компьютерного зрения с десятилетним опытом;
- спикер на образовательных и научно-популярных мероприятиях и конференциях.



## Алексей Мастов

Deep Learning инженер в NVIDIA;

### О преподавателе

- специалист в области компьютерного зрения и машинного обучения;
- участник в проектах по детектированию объектов, 3D SLAM, оценке качества изображений;
- стажировался в группе компьютерного зрения и геометрии ETH Zurich;
- ведущий семинаров и лекции по машинному обучению на разных площадках.

# Особенности программы

- **Специализация во время обучения**

Вы можете выбрать направление, которое вам ближе, и погрузиться в него с головой.

- **Реальные данные**

Для решения задач мы даем настоящие данные, которые максимально приближены к жизненным ситуациям. После курса у вас будет год реального опыта с большим массивом данных.

- **Сообщество, помощь преподавателя и обратная связь**

Вы получите доступ к закрытому сообществу, где сможете обмениваться опытом со студентами. Во время обучения вы будете общаться с преподавателем, который поможет разобраться в задачах и ответит на вопросы.

- **Учителя — практики**

Преподаватели курса — знатоки в своих областях: разработчики и инженеры с опытом работы 10–15 лет.

- **Помощь в трудоустройстве**

Наша цель — помочь освоиться в мире IT и найти работу, которая отвечает вашим компетенциям.

Skillbox

# **Ваше резюме по итогам прохождения курса**





## Иван Иванович

Специалист по анализу  
данных

**100 000 ₽**

ожидаемая зарплата

### Профессиональные навыки:

- Владение языком программирования Python для анализа данных и применения алгоритмов машинного обучения.
- Работа с различными источниками данных: CSV, XML и XLS.
- Знание основ SQL и NoSQL баз данных.

### Навыки в специализации «Анализ данных»:

- R и знание основных библиотек.
- Организация и проведение A/B-тестирования
- Создание аналитических панелей (фреймворки Dash и Shiny)

### Владение инструментами:



R

pandas

Pandas



Jupyter



Numpy



MongoDB



Scipy



Python



Shiny



Dash



Scikit-learn



Matplotlib



SQL



## Максим Максимович

Специалист по  
машинному обучению

**120 000 ₽**

ожидаемая зарплата

### Профессиональные навыки:

- Владение языком программирования Python для анализа данных и применения алгоритмов машинного обучения.
- Работа с различными источниками данных: CSV, XML и XLS.
- Знание основ SQL и NoSQL баз данных.

### Навыки в специализации «Машинное обучение»:

- Разработка рекомендательных систем.
- Знания в областях Machine Learning, Deep Learning: Computer Vision, Image Recognition с использованием библиотек Keras и Tensorflow.
- Поддержка и сопровождение моделей машинного обучения.

### Владение инструментами:



TensorFlow

pandas

Pandas



Jupyter



Numpy



MongoDB



Scipy



Python



SQL



Keras



Scikit-learn



Matplotlib

Skillbox

**Спасибо**