

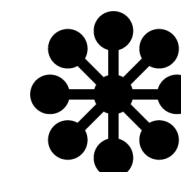
The background is a dark blue gradient. In the center, there is a complex network diagram with red and teal lines connecting various nodes (circles and squares). Some nodes are highlighted with larger circles. The diagram is surrounded by faint, larger squares and lines. Two large, thick blue curly braces are positioned on the left and right sides of the central text, framing it.

# Как войти в сферу работы с данными и стать Data Scientist: пошаговый план





# Вдохновитесь! Да, это важно



Работать с данными сложно. Нужно по-настоящему любить логику и числа, чтобы во всём разобраться. Поэтому для вдохновения посмотрите, насколько востребованы специалисты этого направления.



Уже в 2018 году на одном только рынке США  
**не хватало 200 тыс.  
аналитиков и Data Scientists**  
и 1,5 млн. менеджеров, способных задавать  
правильные вопросы.

✧ по данным McKinsey



Data Scientist — лучшая работа в Америке  
в 2018 году со средней базовой зарплатой  
**\$110 000 в год**

✧ по данным Glassdoor



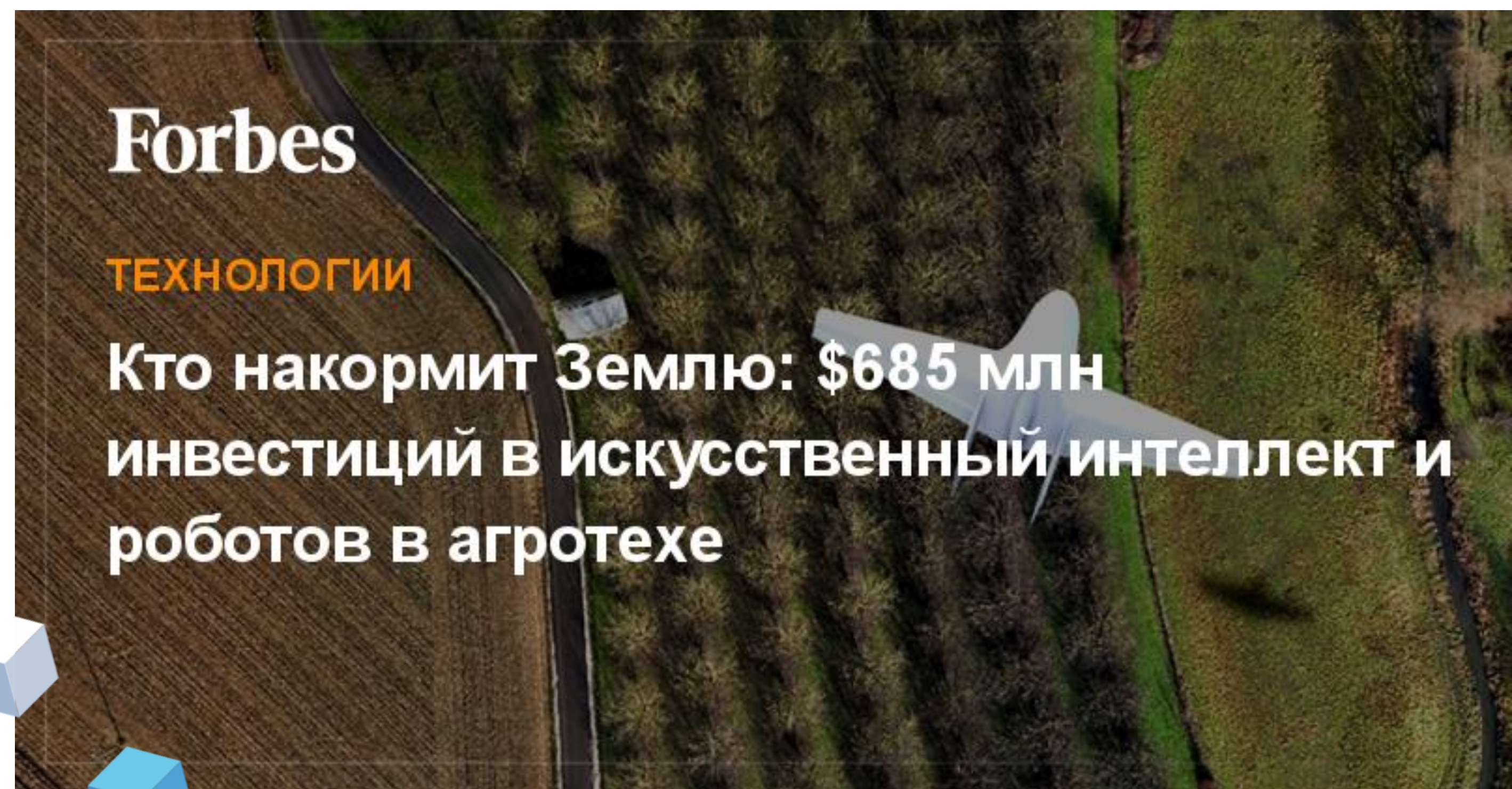


# Специалисты востребованы везде, даже в не самых очевидных областях

## АГРОТЕХ

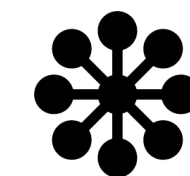
- сбор урожая роботами
- электронная карта полей
- погода
- спутниковые снимки
- техкарты
- управление ресурсами
- скаутинг

По данным экспертов CB Insights, сельскохозяйственные технологические стартапы с 2012 года привлекли более \$800 млн. С 2014 года отмечается устойчивый фокус инвесторов на стартапы, применяющие в сельском хозяйстве робототехнику и машинное обучение. Параллельно с агротехническим сектором эти технологии применяются в здравоохранении, финансах и торговле и других областях, — [Forbes](#).





# Специалисты востребованы везде, даже в не самых очевидных областях



## РИТЕЙЛ

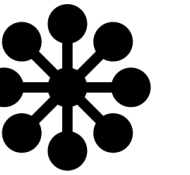
- рекомендательные системы
- программы лояльности
- анализ раскладки товара
- сигнализация о состоянии запасов
- и многое другое.

Роботы, которые занимаются транспортировкой заказов на складах компании Amazon оказались в 4-5 раз эффективнее сотрудников компании, работающих на складе. Те операции, на которые человек тратит 60-75 минут, робот выполняет за 15, — [N+1](#).





# Специалисты востребованы везде, даже в не самых очевидных областях



## КИНОПРОКАТ И ПРОИЗВОДСТВО

- алгоритмы оценки потенциала фильма в прокате
- пок кадровый анализ сцен и сюжетных линий по вкусам аудитории
- чистовой монтаж

В 2012 году в Кино Mail.Ru впервые запустили систему рекомендаций, а затем усовершенствовали её с помощью нейросетей — [кейс на vc.ru](https://vc.ru/1111111).

Карта мирового кинематографа

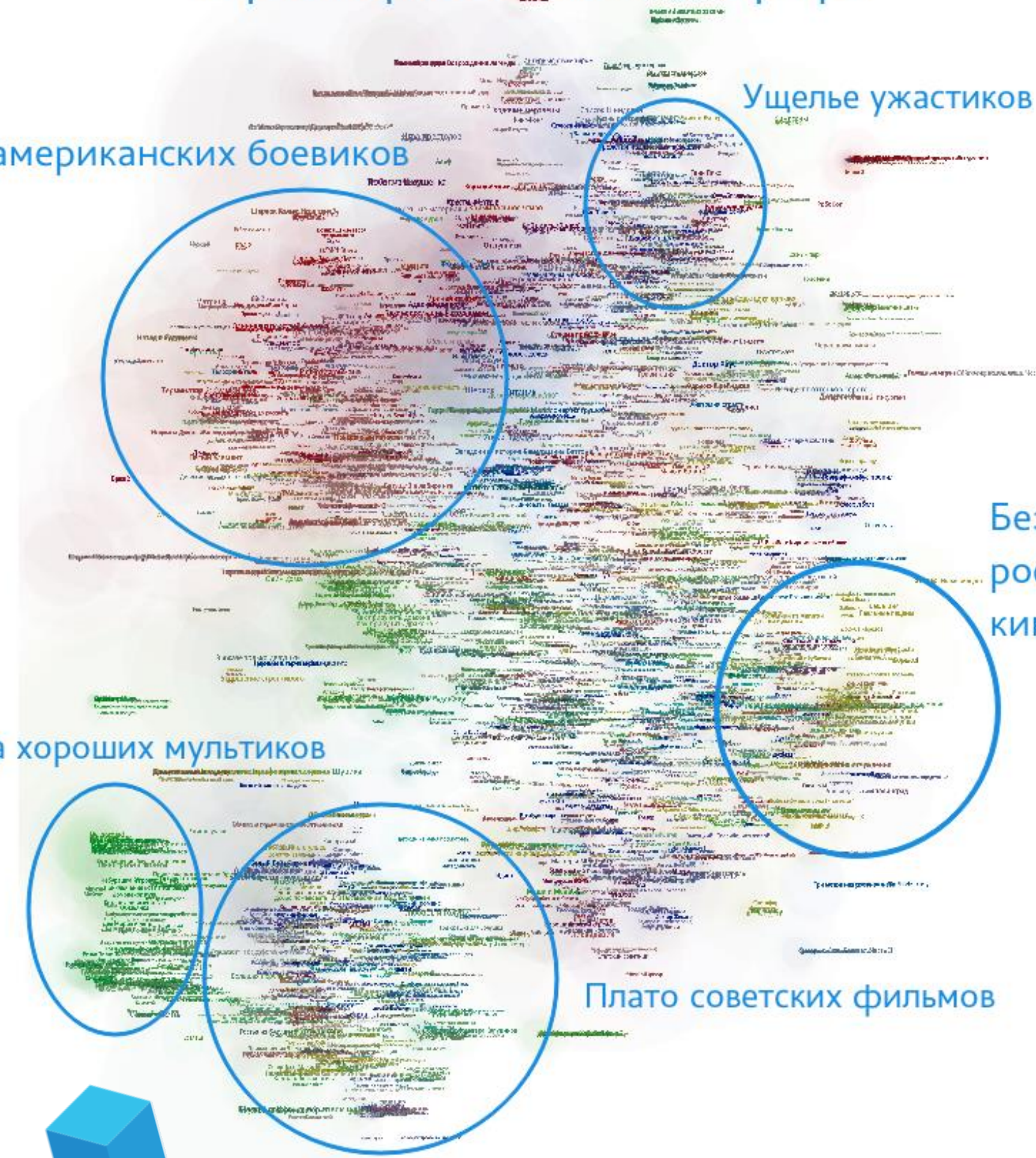
Хребет американских боевиков

Ущелье ужасиков

Бездна  
русского  
кино

Долина хороших мультков

Плато советских фильмов





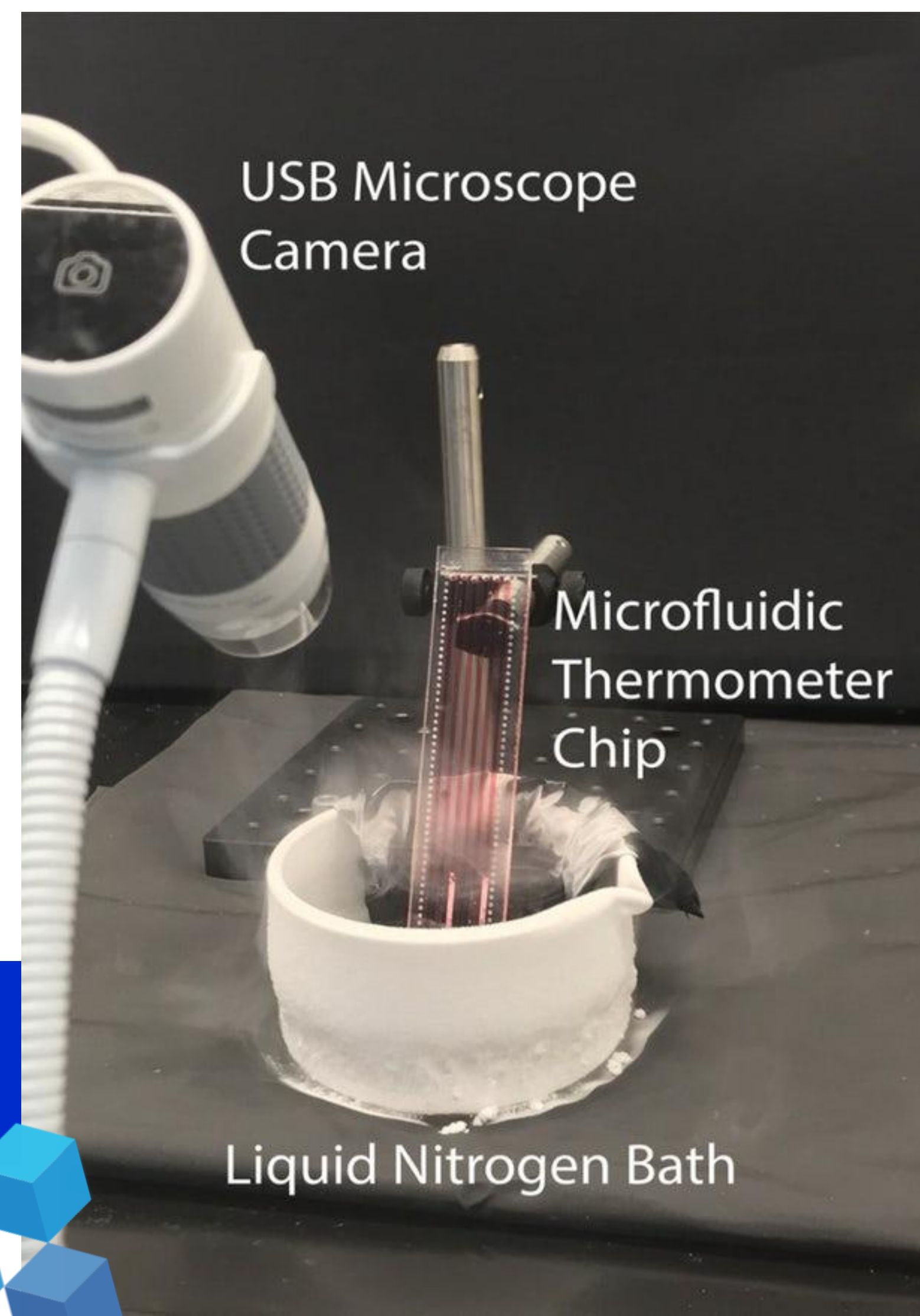
# Специалисты востребованы везде, даже в не самых очевидных областях

## МЕДИЦИНА

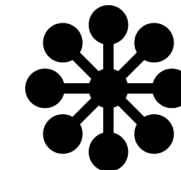
- анализ медицинских карт и прогнозирование вероятных болезней
- подробный анализ состояния здоровья с учетом тысячи показателей, которые не рассматриваются врачами при обычном приеме

ИИ научился определять поддельные лекарства на основе хронопринтов, — [Хайтек](#).

**Данные — новая нефть**



# Изучите терминологию



На всякий случай познакомимся с основными понятиями из сферы Data Science. Мы не будем уходить в дебри и дадим только те термины, которые понадобятся для изучения нашего гайда.

## Data Science

наука о данных, которая анализирует и придает смысл данным, чтобы на их основе принимать оптимальные решения. Включает в себя несколько дисциплин: визуализация, статистика, математика

## Big Data

набор подходов, с помощью которых извлекаются неструктурированные данные

## BI-системы

компьютерные методы и инструменты для бизнеса, которые помогают перевести данные в человекочитаемую форму, пригодную для бизнес-анализа

## SQL

язык запросов, который позволяет получать данные из баз данных, фильтровать, агрегировать, а также импортировать и экспортировать их

## Python

язык программирования, который часто используется для обработки данных и создания моделей анализа

## Dashboard

инструмент для визуализации и анализа информации о бизнес-процессах и их эффективности

## Искусственный интеллект (AI)

способность машин решать творческие задачи. Перед технологиями искусственного интеллекта стоит задача «научить машины думать»

## Машинное обучение

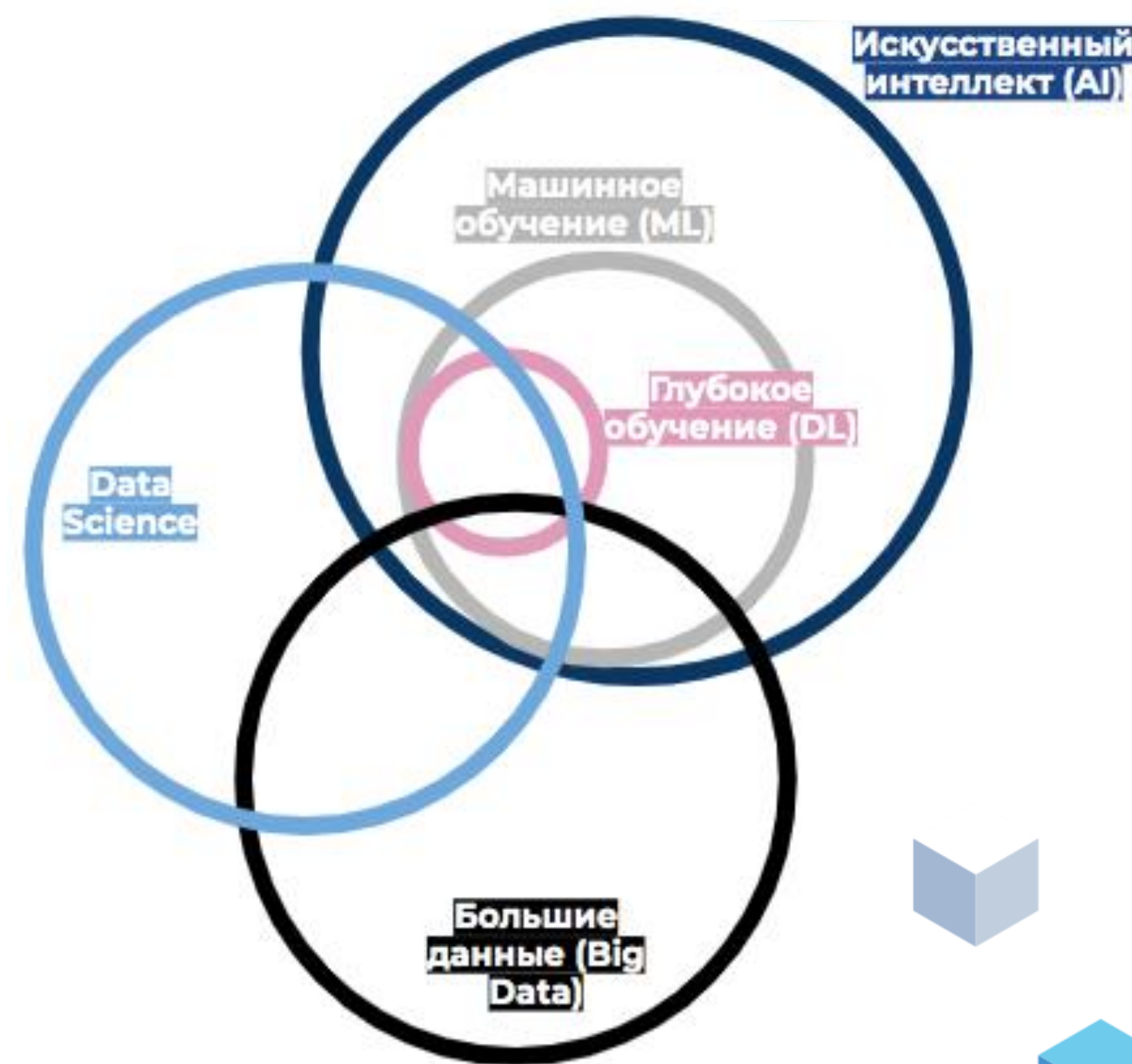
раздел искусственного интеллекта, изучающий инструменты для извлечения инсайтов из данных

## Нейросети

один из видов машинного обучения. Программа работает по принципу человеческого мозга: прогоняет входные данные через систему «нейронов» — более простых программ, взаимодействующих между собой, после чего выдаёт некий результат вычислений на основе этого взаимодействия

## Глубокое обучение (DL)

один из подходов к построению и обучению многослойных нейронных сетей





# Изучите список возможных задач

Новичкам важно понять, какие бизнес-задачи обычно стоят перед специалистами сферы Data Science. Мы будем говорить человеческим языком, чтобы поняли даже начинающие. Если вы уже немного в теме — простите нам упрощения.

«Для начала мудрость аналитика:

Само по себе наличие больших массивов данных не имеет значения, если не будет использовано для получения конкурентного преимущества.»

## Примеры задач, которые решают аналитики и Data Scientist

Руководство интернет-магазина хочет, чтобы вместо одного товара клиенты покупали сразу несколько. Нужна рекомендательная система, которая будет предлагать купить дополнительный товар с учетом характеристик покупателей.

### РЕШЕНИЕ:

Построить «машину», которая сможет выявить закономерности в поведении покупателей и на их основе рекомендовать подходящие товары

Банку нужно определять подозрительные транзакции и останавливать их. Каждую секунду проходит миллион транзакций — отслеживать их вручную просто невозможно.

### РЕШЕНИЕ:

Создать алгоритм, который проанализирует транзакционные активности, изучит признаки подозрительных транзакций и на основе этих данных классифицирует новые транзакции

Врачи не всегда могут точно определить, стоит ли назначать больному операцию: улучшит ли она его состояние, справится ли организм.

### РЕШЕНИЕ:

Написать алгоритм, который спрогнозирует состояние пациента на основе анализа тысяч медицинских карт других пациентов.

Социальной сети нужно организовать ленту пользователя максимально эффективным способом: контента становится слишком много и важно давать пользователю только то, что ему интересно.

### РЕШЕНИЕ:

Создать алгоритм, который сначала определит, что интересно человеку, а затем по этим критериям сформирует для него «умную ленту»

## Остальные вещи, которые уже кажутся привычными:

Спам-фильтры в почте, группировка фото по лицам в iPhone, ранжирование и выдача сайтов в поиске по ключевому запросу, прогноз стоимости ценных бумаг, приложения, распознающие рукописные цифры и буквы, и многое другое — тоже дело рук дата сайентистов.



# Изучите популярные карьерные направления

В сферу трудно попасть абсолютным новичкам. Поэтому лучше всего выбирать направления развития с учетом текущего опыта и места работы. Ниже мы собрали самые частые ситуации и указали средние значения зарплат junior-специалистов по каждому направлению.

Для справки

junior – начинающий уровень  
(0,5-1,5 года опыта)

Следующие уровни:

- middle (1-3 года)
- senior (4-6 лет)
- lead (руководящая позиция)

Ограничения по опыту очень условны: уровень зависит от навыков, насмотренности и компетенций.

Junior – решает задачи под присмотром, middle – решает сам, senior – решает их самым оптимальным способом или руководит решением.

Бэкграунд  
сейчас

**Бухгалтер/  
Финансист/Треjder  
/Экономист**

**Веб-аналитик/  
Маркетолог/  
Sales-менеджер**

**Разработчик/  
Математик/  
Аналитик**

Текущие  
навыки

- Склонность к автоматизации рутинной работы.
- Умение работать с числами.
- Опыт работы на стороне бизнеса.
- Базовое знание Excel.

- Опыт работы в digital.
- Базовое знание Google Analytics и Яндекс Метрики.
- Быстрое освоение новых инструментов.

- Исследовательский интерес.
- Профильный технический опыт обучения или работы.

Куда лучше  
расти

**BI-аналитик**

Средняя з/п в Москве – 80 000 ₽

Средняя з/п в регионах – 35 000 ₽

**Аналитик данных**

Средняя з/п в Москве – 90 000 ₽

Средняя з/п в регионах – 40 000 ₽

**Data Scientist**

Средняя з/п в Москве – 113 000 ₽

Средняя з/п в регионах – 50 000 ₽

Сколько времени  
понадобится  
на обучение

~ 2 года

~ 2 года

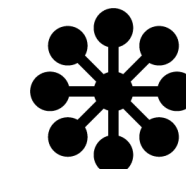
~ 2 года

Советуем посмотреть запись  
бесплатного занятия:

- [Data Science для начинающих: обзор сферы и профессий](#)
- [Рынок труда аналитиков и Data Scientists](#)
- [Вебинар «Data Science для начинающих: обзор сферы и профессий»](#)



# Освойте инструменты, необходимые junior-специалисту



Не переживайте, если не вполне понимаете, чем отличается BI-аналитик от аналитика данных или Data Scientist. Описали задачи и навыки junior-специалистов по каждому из направлений.

## BI-аналитик

Собирает исходные данные и визуализирует их так, чтобы бизнес мог принимать на основе собранных дашбордов оптимальные решения. В своей работе использует готовые инструменты визуализации вроде Power BI, Docker, Power Query и др. Важно: ему не нужно уметь программировать. Единственный нужный ему язык – это язык запросов SQL, с помощью которого специалист обращается к базе данных и выгружает, сортирует и обрабатывает имеющиеся там данные. Потребуется знать только набор определенных команд — что не так сложно.

**Пример:** руководство интернет-магазина хочет видеть, как у него дела с трафиком и конверсией в заявку. BI-аналитик с помощью готового инструмента собирает дашборд, на котором собственники могут в легкочитаемой форме видеть эти показатели и быстро замечать, где проблема.

### Ключевые слова

визуализация

отчетность

### Инструменты

- Power BI,
- SQL,
- Hadoop,
- Power Query,
- Электронные таблицы
- Инструменты доступа к данным и хранилища данных
- Инструменты визуализации данных и построения отчетов OLAP-инструменты
- Инструменты дата-майнинга анализа данных: SAS, R, Weka, Knime, RapidMiner.

### Hard Skills

- Разработка требований к аналитическим инструментам, которые создают программисты, и сопровождение их внедрения в компании.
- Подготовка аналитических материалов для выработки стратегических решений.
- Поддержка актуальности и развитие системы управленческой отчетности.

### Soft Skills

- Скетчинг и прототипирование.
- Развитый эмоциональный интеллект.
- Умение перевести требования бизнеса в конкретные задачи для других участников проекта.

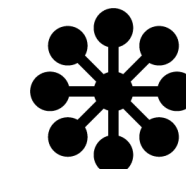
### Где бесплатно

#### почитать или посмотреть:

- [Вебинар «Полезные навыки работы в Excel, о которых не знают профи»](#)
- [12 книг по Data Science для новичков и продвинутых](#)
- [Вебинар «Просто о сложном: Data Science, Machine Learning и Big Data»](#)



# Освойте инструменты, необходимые junior-специалисту



Не переживайте, если не вполне понимаете, чем отличается BI-аналитик от аналитика данных или Data Scientist. Описали задачи и навыки junior-специалистов по каждому из направлений.

## Аналитик данных

Также собирает исходные данные, но не только визуализирует для бизнеса, но и сам анализирует их, чтобы решить текущие проблемы компании. Для этого он выдвигает и тестирует гипотезы, проводит статистические тесты. Еще одно отличие от BI-аналитика: использует не только готовые инструменты вроде Power BI, но и умеет работать с языками программирования. Работает с гипотезами, а не аналитикой ради отчетов.

**Пример:** руководство интернет-магазина видит, что конверсия в заявку снизилась, при том что никаких существенных изменений в функциональность сайта не вносилось. Аналитик данных собирает данные, на основе этого выдвигает ряд гипотез и начинает их тестировать, чтобы решить текущую проблему бизнеса.

### Ключевые слова

принятие решений

текущие проблемы

моделирование

### Инструменты

- Pandas
- Hadoop
- NoSQL
- Python: Scikit-learn, numpy, scipy, plotly, pandas и прочие библиотеки около анализа данных
- Docker
- Электронные таблицы
- Инструменты доступа к данным и хранилища данных
- Инструменты визуализации данных и построения отчетов OLAP-инструменты
- Инструменты дата-майнинга анализа данных: SAS, Python, R, Weka, Knime, RapidMiner.

### Hard Skills

- Получение, очистка, преобразование, визуализация и представление данных.
- Сбор и анализ требований заказчиков к отчетности.
- Знание ключевых математических методов и основ статистики.
- Умение работать с языками программирования.

### Soft Skills

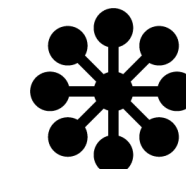
- Умение видеть за метриками физический смысл и находить причины и взаимосвязи событий.
- Способность давать рекомендации бизнесу.
- Умение мыслить абстрактно.

Где бесплатно  
почитать или посмотреть:

- [Вебинар «Аналитика данных для новичков: обзор лучших практик»](#)



# Освойте инструменты, необходимые junior-специалисту



Не переживайте, если не вполне понимаете, чем отличается BI-аналитик от аналитика данных или Data Scientist. Описали задачи и навыки junior-специалистов по каждому из направлений.

## Data Scientist

Data Scientist также преобразуют сырые данные в полезную информацию, но не для ответа на текущие вопросы компании, а для принятия будущих стратегических решений. Этот специалист умеет задать важные вопросы. Как и аналитик данных, использует не только готовые инструменты, но и умеет работать с языками программирования.

**Пример:** руководство интернет-магазина знает, что за десятки лет существования они накопили большое количество данных о своих клиентах. Но не используют их для получения какого-либо конкурентного преимущества. Они ставят, можно сказать, творческую задачу дата сайентисту: собрать все имеющиеся данные, выявить закономерности и извлечь из них пользу для будущих решений. Казалось бы: пойдя туда, не знаю куда? Но в этом и есть работа дата сайентиста — самому выявить точки роста и с помощью данных повлиять на стратегию компании.

### Ключевые слова

стратегия

правильные вопросы

понимание

### Инструменты

- OpenCV
- PostgreSQL
- NLTK
- Python: Scikit-learn, numpy, scipy, plotly, pandas, Matplotlib, Seaborn и прочие библиотеки около анализа данных
- Tensorflow
- Инструменты доступа и обработки данных (СУБД, SQL, Хранилища данных, ETL)
- Инструменты BI (Tableau, Power BI)
- Специализированные инструменты Data Science (R, Python, SAS, Alteryx, Weka)
- Интерактивные оболочки, ноутбуки (Zeppelin, Jupiter)
- Инструменты автоматизации и развёртывания (Docker, Airflow)

### Hard Skills

- Поиск трендов и закономерностей в данных.
- Определение необходимого набора данных для решения задач.
- Сбор, предобработка, очистка и вариация необходимых данных из различных источников.
- Разработка и применение моделей и алгоритмов анализа данных.
- Программирование (R, SAS, Python).
- Преобразование и подготовка данных (SQL, Python).
- Исследование данных (exploratory data analysis).
- Статистический анализ (statistical inference).
- Регрессионный анализ (regression analysis).
- Машинное обучение (machine learning).

### Soft Skills

- Идентификация значимых аналитических проблем и задач.
- Интерпретирование результатов анализа, поиск решений проблем и новых возможностей.
- Представление результатов заказчику.

### Где бесплатно

почитать или посмотреть:

- [Гайд в мир Data Science для начинающих](#)
- [10 типов структур данных, которые нужно знать + видео и упражнения](#)
- [12 книг по Data Science для новичков и продвинутых](#)



# Используйте бесплатные источники информации

Для специалиста по работе с данными важно постоянно прокачиваться, следить за трендами индустрии. Выше мы уже дали ссылки на статьи и вебинары, с помощью которых можно сориентироваться в сфере и понять, куда двигаться дальше, добавим к ним еще несколько источников информации — русско- и англоязычных.

## Русскоязычные источники

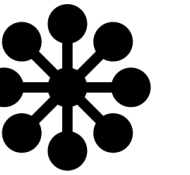
- [DeepLearning](#)  
новости и материалы из области Deep Learning (Глубокое обучение) с упором на Computer Vision (Машинное зрение) и русскоязычную информацию
- [Новости Науки и Технологий](#)  
научно-популярный youtube-канал о значимых достижениях, изобретениях науки
- [Хайтек](#)  
медиа про высокие технологии в России

## Англоязычные источники

- [www.dataschool.io](http://www.dataschool.io)  
портал с туториалами для начинающих
- [t.me/opendatascience](https://t.me/opendatascience)  
телеграм-канал с новостями об AI, больших данных, машинном обучении, статистике
- [www.analyticsvidhya.com](http://www.analyticsvidhya.com)  
портал о Data Science с форумом, полезными статьями и вакансиями



# Правильно составьте резюме



Данные — самый ценный на рынке продукт, и вакансий всё больше и больше. Но компаниям нужны не просто любители анализа и исследований — нужны те, кто умеет решать задачи реального бизнеса. Поэтому и резюме важно составлять с ориентацией на результат.

## Признаки хорошего резюме

Список технологий, в которых вы действительно разбираетесь. Не нужно распыляться и писать обо всем, что немного поизучали. Еще дилетанта может выдать структура резюме: когда знание офисных программ и профессиональных инструментов аналитика просто идут одним списком — располагайте технологии так, как они связаны на рынке.

### Неправильно

Знаю: SQL, Power BI, Python, Excel, Feature Engineering, Jupyter, Tensor Flow, Битрикс.

### Правильно

В совершенстве владею: SQL, Power BI, Excel.  
Немного знаком: Feature Engineering, Jupyter.

Описание опыта через достижения, решенные задачи, интересные технические подробности. Укажите, какую часть проекта разработали именно вы.

### Неправильно

Работал с данными, осуществлял анализ — и другие общие слова.

### Правильно

Разработал алгоритмы предсказания аномалий с использованием ML, что сократило операционные расходы компании в 1,5 раза. Внедрил алгоритм семантического анализа.

Ссылка на портфолио с подробным описанием ваших проектов и выводов по ним.

### Неправильно

Много схем, мало выводов.

### Правильно

Портфолио на LinkedIn, Kaggle или GitHub с наиболее сложными проектами, которые описаны следующим образом: как начался проект, как вы оценили скоп задач, какие цели поставили, чем массив данных был интересен, и обязательно остановитесь на результате — чего вы добились работой над проектом, достигли ли цели.

### Также читайте

- [Резюме и портфолио для аналитика и data scientist: советы HR-специалистов](#)
- [Кто такой Data Scientist – глазами работодателя](#)
- [Интервью с Авито и Spice IT о роли Data Scientist](#)

« Хорошее резюме не должно быть больше 1-2 страниц. Откликаясь на вакансию, напишите сопроводительное письмо: почему вам интересна вакансия, почему ваш опыт релевантен, любимые профессиональные книги и ресурсы //



# Заводите знакомства в индустрии

Сейчас все ищут кадры в социальных сетях и через знакомых. И чем больше разработчиков в вашем списке друзей, тем выше вероятность найти крутую вакансию. Посещайте конференции, митапы и хакатоны. Вступайте в профильные группы, регистрируйтесь на форумах и программерских сайтах.

## Вот несколько полезных ссылок

---

### Форумы

- [www.kdnuggets.com](http://www.kdnuggets.com)  
портал по интеллектуальному анализу данных
- [mloss.org/software](http://mloss.org/software)  
коллективный сайт разработчиков  
открытого софта для машинного обучения
- [www.gotai.net](http://www.gotai.net)  
русскоязычный портал с полезными  
материалами и форумом для профессионалов
- [www.kaggle.com](http://www.kaggle.com)  
еще один портал для аналитиков, на котором  
они выкладывают свои открытые проекты

### Telegram-чаты и группы в соцсетях

- [vk.com/weirdreparametrizationtrick](https://vk.com/weirdreparametrizationtrick)  
группа с профессиональными шутками  
и мемами, в которой наверняка сидят  
люди из сферы
- [t.me/@opendatarussiachat](https://t.me/@opendatarussiachat)  
чат тех, кто занимается  
открытыми данными в России

Регистрируйтесь на  
регулярный бесплатный митап  
Нетологии [«Карьера в Data  
Science для начинающих»](#).  
Если вы живете в Москве —  
приходите лично. Если в  
регионе — смотрите онлайн-  
трансляцию.



# Погрузитесь в сферу совместно с экспертами

Разобраться в сфере Data Science и начать в ней работать можно и самостоятельно, но это займет слишком много времени — а ведь она так стремительно меняется!

Сэкономить время и сориентироваться в области работы с данными всего за три недели поможет обзорный курс «Старт в Data Science». Вы оцените, насколько вам интересна работа с данными, и поймете, куда развиваться дальше.

## 7 занятий

с практикующими  
экспертами, которые  
расскажут просто о сложном

## 5 инструментов

освоим must have-набор  
любого специалиста по  
работе с данными

## Теория + практика

все полученные знания  
отработаем на реальных  
кейсах и инструментах

## Развитие карьеры

поможем в составлении  
траектории обучения и  
выборе специализации

Запишитесь на курс и сделайте первый шаг в работе с данными!

[ИЗУЧИТЬ ПРОГРАММУ](#) →