

Data science

# Лекция 1. Применение Data science

2025/2026 учебный год

Доцент кафедры МО&МО, Махно В.В.



Кодовое слово для способа МАГИСТРЫ 2026 = 12022026

## Data Science и инструменты Анализа Данных



# Что включает в себя Data Science:

**Data Science** — это область, которая помогает получать ценные знания из данных с помощью программирования, статистики и машинного обучения.

**сырые данные → анализ → выводы → реальные задачи.**

**Data Science** включает:

- **Сбор данных** — из таблиц, сайтов, сенсоров, API
- **Очистка и подготовка данных** — чтобы данные стали пригодными для анализа
- **Анализ данных (EDA)** — визуализации, группировки, поиск закономерностей
- **Моделирование** — обучение моделей для предсказаний
- **Интерпретация** — объяснение результата людям или бизнесу

# Data Science используется:

- Рекомендательные системы (Netflix, Spotify)
- Финансовые модели (одобрение кредита, прогноз акций)
- Маркетинг (анализ поведения клиентов)
- Здравоохранение (диагностика по снимкам)
- Спорт (оценка игроков, стратегия)

# Кто такой Data Scientist?

Это человек, который может:

- понять бизнес-проблему,
- собрать и проанализировать данные,
- построить модель,
- и **предложить решение**, основанное на данных.

**Итак:**

**Data Science — это мост между данными и реальными решениями.**

**Это профессия, где сочетаются математика, код, логика и креатив.**

# Применение Data Science:

## YouTube, TikTok, Spotify

Алгоритмы рекомендуют видео и музыку на основе:

- Истории просмотров
- Поведения пользователей, похожих на вас
- Времени суток, дня недели и других факторов

→ Это **рекомендательные системы**, они основаны на анализе больших данных и машинном обучении.

## Онлайн-магазины: Wildberries, Ozon, Amazon

Data Science помогает:

- Показывать товары, которые вас заинтересуют
- Прогнозировать спрос
- Оптимизировать логистику

→ Пример: если вы ищете «кроссовки», система предложит похожие товары, основываясь на поведении других покупателей.

## Банки и финансы

Используют модели, чтобы:

- Выявлять мошенничество
- Предсказывать вероятность возврата кредита
- Предлагать персональные продукты

→ **Кредитный скоринг** — классическая задача классификации.

# Применение Data Science:

## **Здравоохранение**

Data Science применяют для:

- Распознавания болезней на снимках (рентген, МРТ)
- Прогнозирования рисков заболеваний
- Персонализированных рекомендаций по лечению

## **Спорт и аналитика**

Клубы анализируют:

- Игру каждого спортсмена
- Оптимальную тактику
- Кто из игроков «перспективен» для покупки

## **Промышленность, транспорт, логистика**

- Предсказание поломок оборудования
- Оптимизация маршрутов доставки
- Управление складом и запасами

*Data Science используется везде, где есть данные. А данные — повсюду: в интернете, в банках, в медицине, в бизнесе, в телефоне каждого из нас.*

# Обзор профессий в Data Science

## Аналитик данных (Data Analyst)

### Что делает:

- Извлекает данные из баз
- Делает отчёты, дашборды, визуализации
- Отвечает на вопросы бизнеса: «почему продажи упали?», «кто наш клиент?»

### Навыки:

- SQL, Excel, Python (Pandas, Matplotlib)
- Tableau / Power BI
- Бизнес-мышление

**Цель:** объяснить прошлое и настоящее.

# Обзор профессий в Data Science

## Data Scientist

### Что делает:

- Разрабатывает модели прогнозирования
- Решает более сложные задачи: рекомендации, классификация, прогноз
- Работает с данными глубже, часто строит end-to-end решения

### Навыки:

- Python (Pandas, Scikit-learn, ML-библиотеки)
- Статистика, математика
- Построение и оценка моделей
- Иногда — немного SQL, визуализации

**Цель:** предсказать будущее и автоматизировать принятие решений.

# Обзор профессий в Data Science

## ML-инженер (Machine Learning Engineer)

### Что делает:

- Берёт модель от Data Scientist и делает так, чтобы она **работала в проде**
- Занимается инфраструктурой, масштабируемостью, скоростью, API

### Навыки:

- Python (или другой язык), Git, Docker, CI/CD
- MLflow, FastAPI
- Глубокое понимание пайплайна моделей

**Цель:** запустить модель в реальном приложении — стабильно и быстро.

# Обзор профессий в Data Science

## Дополнительно

- **Research Scientist** — занимается новыми алгоритмами, работает в больших лабораториях
- **Data Engineer** — строит системы для сбора, хранения и обработки данных

## Как выбрать

Графики и отчёты	Data Analyst
Математику и модели	Data Scientist
Кодить и автоматизировать	ML Engineer

# «Данные — это топливо 21 века» (2006 г). Без них не работают рекомендательные системы, не строится логистика, не принимаются бизнес-решения.

- Компании ежедневно генерируют терабайты данных — о клиентах, продажах, процессах.
  - «Сырые данные» — бесполезны. Нужны специалисты, кто может **выжать из них пользу**.
  - Поэтому **спрос на Data Science растёт во всём мире** — от стартапов до крупнейших корпораций.
- ❓ Прогноз McKinsey: к 2030 году нехватка специалистов по данным в мире превысит **250 000 человек**.
- ❓ Вывод: умение работать с данными = навык будущего, как раньше умение пользоваться компьютером.

Посмотреть на hh.ru «Data Scientist» и «Data Analyst»

# Востребованность профессий

hh.ru → “Data Scientist” — в крупных городах это **сотни вакансий**.

• Для «Data Analyst» — ещё больше: от ритейла до финтеха.

• **LinkedIn** показывает: профессии, связанные с данными, — в топ-10 самых быстрорастущих в мире.

📌 Зарплаты (по данным hh.ru, весна 2025):

зарплаты зависят от города, компании и реальных навыков — чем лучше портфолио и понимание, тем выше предложение.

Профессия	Джун (₽)	Мидл (₽)	Сеньор (₽)
Data Analyst	80–120 тыс.	130–180 тыс.	200–250 тыс.
Data Scientist	100–150 тыс.	180–250 тыс.	300+ тыс.
ML-инженер	120–160 тыс.	200–270 тыс.	350+ тыс.

# Сравнение с другими IT-направлениями

Направление	Что делает	Где применяется	Что особенного
<b>Frontend</b>	Делает сайты красивыми	Веб-разработка	Много визуального, важно внимание к UX
<b>Backend</b>	Обрабатывает логику сайта/сервиса	Сайты, API, базы данных	Много архитектуры, интеграций
<b>QA</b>	Тестирует программы	Везде, где пишется код	Важно внимание к деталям
<b>Data Science</b>	Анализирует данные и строит модели	Финтех, маркетинг, медиа, логистика и др.	Высокий уровень абстракции, автоматизация
<b>ML Engineer</b>	Запускает модели в реальную жизнь	ИИ-продукты, рекомендации, scoring	Сильная инженерия, продакшн, DevOps-навыки

# Данные

- [?] Данные — это основа всего
- Без данных нет Data Science.
- Data Scientist без данных — как повар без продуктов [?].
  
- [?] Что такое данные?
- Данные — это любая зафиксированная информация, которую можно анализировать:
  
- Пример      Что это за данные?
- [?] В интернет-магазине      Название товара, цена, количество покупок
- [?] В медицинской системе      Возраст пациента, диагноз, результаты анализов
- [?] В приложении      Время использования, клики, переходы по экранам
- [?] В Spotify      Треки, лайки, пропуски, плейлисты

Каждый набор данных — это как таблица Excel:

Строки — объекты (пользователи, пациенты, товары)

Столбцы — признаки (возраст, цена, диагноз, рейтинг)



Кодовое слово для способа МАГИСТРЫ 2026 = **12022026**