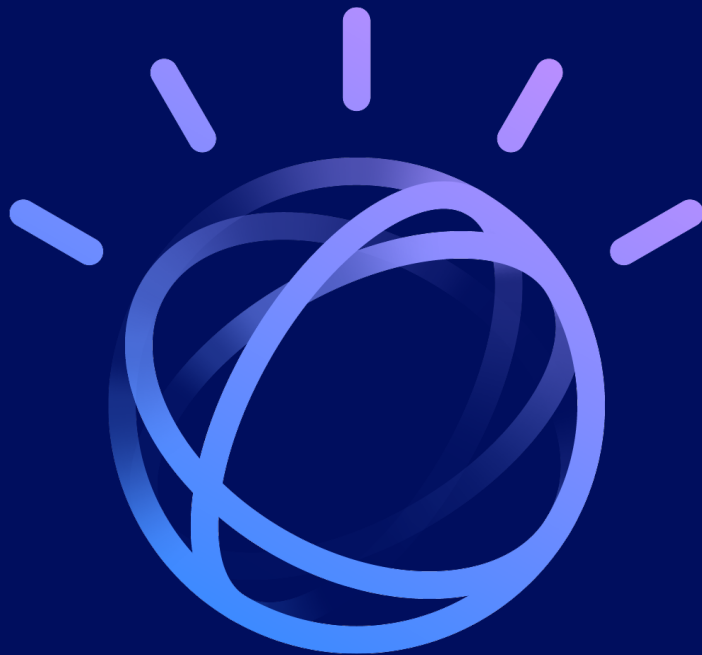


# **Machine Learning – машинное обучение или как сделать сказку былью**



—  
**Roman Shempel**  
**Partner Sales Manager**  
**IBM Cloud, Russia**

# Data Science & Machine Learning: дефиниция

## Machine Learning

это набор моделей из алгоритмов, которые учатся и создают прогнозы на основе подготовленных данных

## Data science

это область исследований данных, объединяет методы обработки данных для получения знаний из данных

## Artificial Intelligence

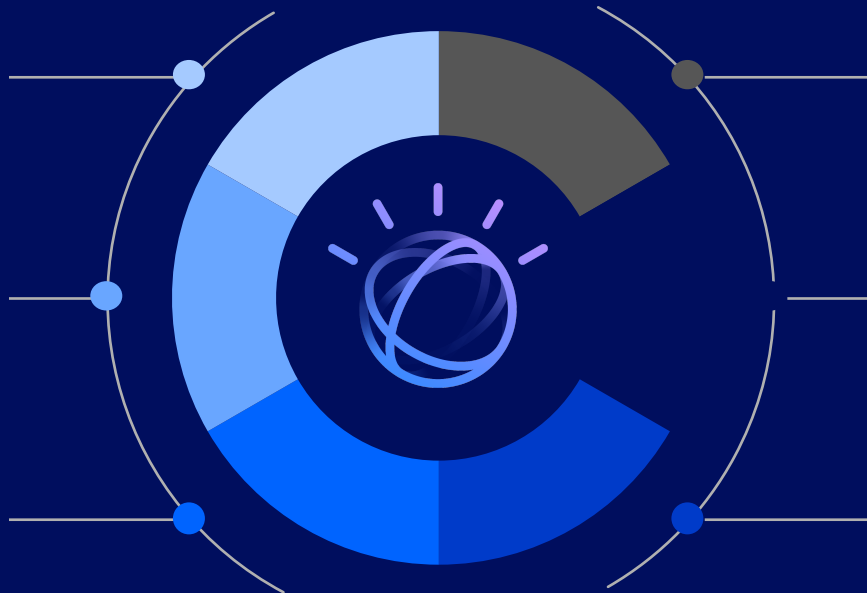
это системы способные дополнять человека на разных этапах принятия решений

# Data Science & Machine Learning: области анализа данных

Распознавание речи,  
лиц и зрительных  
образов

Беспилотные  
механизмы

Интерпретация и  
принятие решений



Выявление спама  
и мошенничества

Рекомендательные  
системы

Интеллектуальные  
базы знаний:  
анализ текста и  
контентная  
аналитика

# Data Science & Machine Learning: академическая среда опережает бизнес

Пример некоммерческого использования **Machine Learning**:  
победа Дональда Трампа и Cambridge Analytics

Бюджеты на компании кандидатов в президенты США:

\$897,7 млн Клинтон

– 15,8 млн on-line реклама – пятая строчка

\$429,5 млн Трамп

– 58,6 млн digital – вторая строчка после общей рекламы



# Data Science & Machine Learning: это процесс

Результаты



## Data Engineer

Решают как организовать работу с данными и обеспечить работу с данными



## Data Scientist

Погружаются глубоко в данные, чтобы извлечь скрытые идеи для бизнеса



## Business Analyst

Работают с данными, чтобы применить идеи к бизнес-стратегии



## Developer

Работают с данными и моделями и пишут код для создания приложений

# Data Science & Machine Learning: это команда



## Business Analyst

Носитель индустриальных знаний, необходимых при построении моделей.

- Понимание бизнес-задачи и данных которые могут быть релевантны для ее решения
- Экспертные знания о факторах влияющих на целевое событие.

**Subject matter expert, content strategist.**



## Data Scientist

Подготовка данных, построение прогнозных моделей и оценка их качества.

- Формализация постановки бизнес-задачи на математическом языке. Тесно взаимодействует с индустриальным экспертом
- Разработка прогнозных моделей. Знание технологий машинного и глубокого обучения, подходов к оценке качества моделей.

**ML/DL engineer, Modeler, Data Miner**



## Data Engineer

Отвечает за организацию данных и их доступность.

- Отвечает за ETL-процесс. Имеет дело со Spark, Hadoop, and HDFS.
- Помогает специалисту по анализу данных внедрить построенную модель в бизнес-процессы.

**Data infrastructure engineer**



## Developer

Разрабатывает приложения в соответствии с бизнес-требованиями.

- Разработка пользовательских интерфейсов, диалоговых окон
- Включение в приложение результатов работы прогнозных моделей

**Front-end, back-end, full stack, mobile or low-code developer**

# Data Science & Machine Learning: проблематика

## Отсутствие единой платформы для инструментов

- Использование единственного подхода
- Ручная сборка всей последовательности шагов
- Не все подходы хорошо/удобно уживаются

## Взаимодействие

- Управление версиями проектов
- Повторное использование уже готовых активов
- Извлечение выводов из результатов работы

## Управление ресурсами

- Масштабирование инфраструктуры
- Диспетчеризация ресурсов



# IBM Watson Studio

Open source инструменты (R, Python, Scala) и дополнения от IBM (SPSS Modeler, Decision optimization, Neural Network Modeler, Data refinery)

Поддержка наиболее распространенных и популярных data science библиотек: Torch, Keras, TensorFlow, Mllib

Возможность выбора инфраструктуры для каждого проекта (CPU\GPU, Storage, Spark etc.)

Создание проектов, поддержка версии и ролевой модели

Разграничение прав доступа





# Watson Studio: автоматизация процесса Data Science & Machine Learning

Подключение и доступ к данным

Подбор релевантных данных

Подготовка данных

Построение прогнозных моделей

Внедрение моделей в бизнес-процессы

Управление жизненным циклом моделей

Подключение к различным источникам данных: текстовым файлам, базам данных, hdfs в облаке или локально

Поиск, а также разработанных ML\DL моделей, Jupiter Notebooks при помощи Knowledge Catalog (с интеллектуальным поиском и разграничением прав доступа между пользователями)

Очистка и подготовка данных с помощью инструмента Data Refinery или стандартных open source библиотек.

Построение моделей с использованием open source инструментов и интерактивного интерфейса продукта SPSS Modeler. Возможность автоматического выбора наилучшей модели.

Простое развёртывание моделей и автоматическое масштабирование для online и batch скоринга

Контроль качества работы моделей в реальном времени, уведомления о падении качества, гибкий подход к дообучению и перестроению моделей

# Data Science & Machine Learning: это работа с данными



**Агрегация – сбор, хранение и обработка данных**



**Интеграция – подключение к источникам данных**



**Подготовка – качество и консистентность данных**

**Управление – создание репозитория метаданных**

**Отчетность – представление и происхождение данных**

# Data Science & Machine Learning: отраслевые кейсы

## Маркетинг

Персонализированные предложения, оптимизация маркетинговых кампаний

## Производство

Прогнозное обслуживание оборудования, оптимизация производства, прогнозирование спроса

## Банковский сектор

Сегментация клиентов, кредитный скоринг, выявление мошенничества

## Транспорт

Управление ценовой политикой, прогнозирование спроса, беспилотные автомобили



## Розничная торговля

Чековая аналитика, удержание клиентов, прогноз продаж, прогнозирование склонности к покупке

## Здравоохранение

Прогнозирование развития заболеваний, анализ дожития, прогноз заболеваемости и спроса на лекарства

## Электроэнергетика

Прогнозирование энергопотребления, оптимальное управление ресурсами

## Безопасность

Выявление аномалий, анализ логов

# IBM

НЕМНОГО КЕЙСОВ



# Kiddicare – UK-магазин товаров для детей и их родителей

# 3

## Задачи

Оптимизация  
маркетинга

Клиентская аналитика

Прогнозирование и  
управление спросом

# 50%

## увеличение числа товаров в заказе

Классические  
упражнения с анализом  
корзины потребления и  
клиентским профилем в  
online-версии  
магазинов

# 20%

## увеличение объема online-заказов

Профилирование  
клиентов и  
предложение  
приобрести вместе с  
выбранным товаром  
дополнительные

# AUCHAN – французкая розница формата гипермаркет

## 2

### Задачи

Таргетированный маркетинг

Формирование индивидуальных продуктовых предложений

## 65%

### Увеличение отклика на маркетинговые кампании

Сегментирование клиентов и таким образом отправка им таргетированных предложений вместо бестолкового бродкаста

## 33%

### Уменьшение стоимости прямого маркетинга на 33% за один месяц

Вместо бомбардировки площадей получился таргет, как результат – сокращение общих затрат на маркетинг в каналах на треть

# Best Buy – крутые в online

# 3

## Задачи

Прогноз отклика  
Клиентская аналитика:  
профилирование и  
сегментация  
Оптимизация  
маркетинговых  
кампаний

# 7%

## Уменьшение затрат на маркетинг

Благодаря тому, что online-бизнес уже предполагал жесткое профилирование клиентов, удалось снизить затраты на маркетинг всего на 7%, что неплохо для оптимизации бизнеса

# x 10

## Увеличение эффективности маркетинга

При этом эффективность (отклик и конверсия) дали эффект в 10 раз

# Redcats – online-мода

2

Задачи

- 
- 

122%

ROI за год

Затраты на ПО + косты на стафф отбились за год, принеся прибыль. При этом за счет верно рассчитанной модели затраты на маркетинг в целом настолько сократились, что привели к быстрой окупаемости всего проекта ML

90%

Уменьшение времени по созданию и управлению маркетинговыми кампаниями



# Thank you

Roman Shempel  
Partner Sales Manager

—

[rshempel@ru.ibm.com](mailto:rshempel@ru.ibm.com)

+7 985 852 56 84

[ibm.com/ru](http://ibm.com/ru)

