# Лекция 3

# «Машинное обучение»

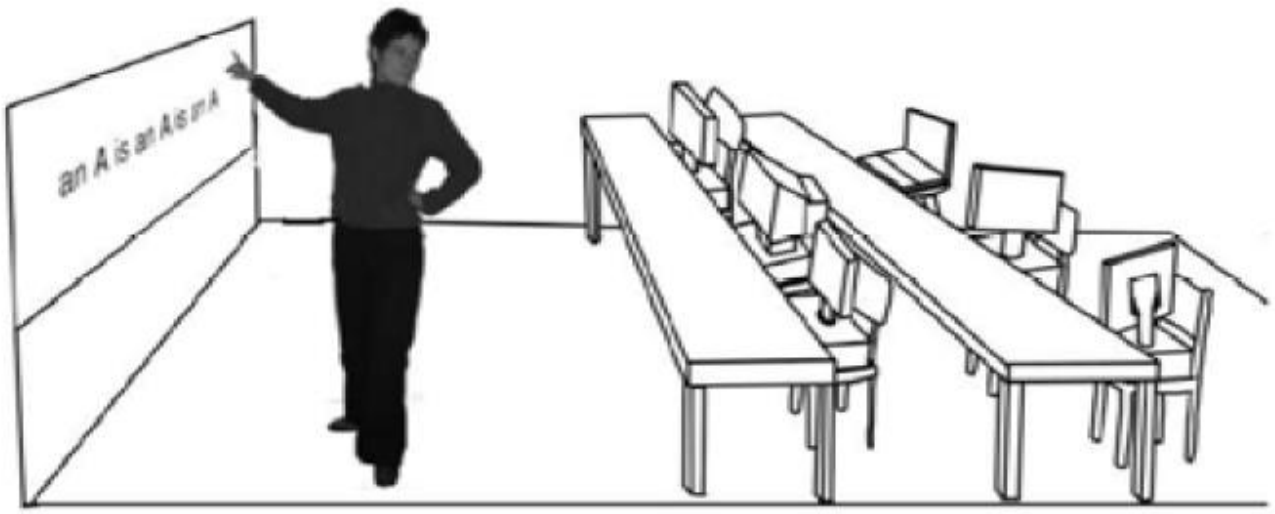
Автор: д.т.н Джомартова Ш.А.

# Обучение vs Интеллект



# Что такое машинное обучение?

- Обучение = приспособление
  - Обучаясь, мы становимся более приспособленными к окружающей среде
- Обучение ≠ «заучивание наизусть»
  - Заучить наизусть – для машины не проблема
  - Мы хотим научить машину делать выводы!

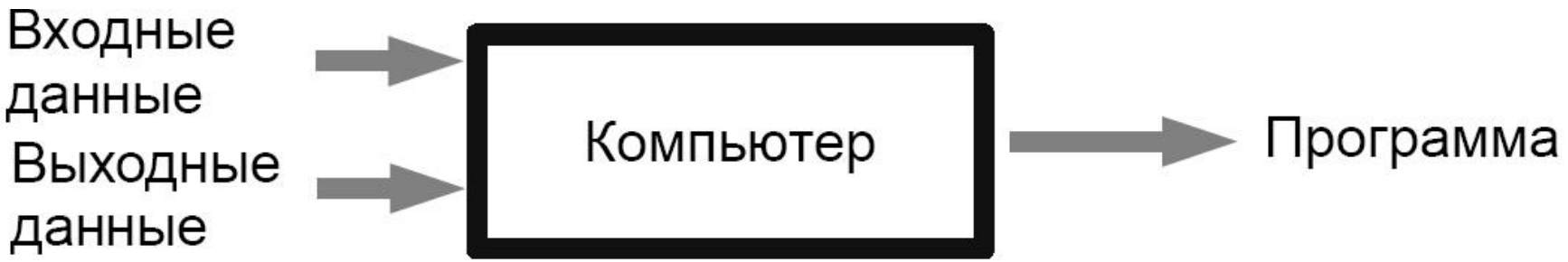


# Машинное обучение. Что это?

- Использование программ




- Создание программ при помощи машинного обучения



# Чем отличаются данные от знаний?

# Данные



Таблицы



Текст



Сигналы



Звук



Изображения



Видео

# Данные

Таблицы



Текст



Сигналы



Звук

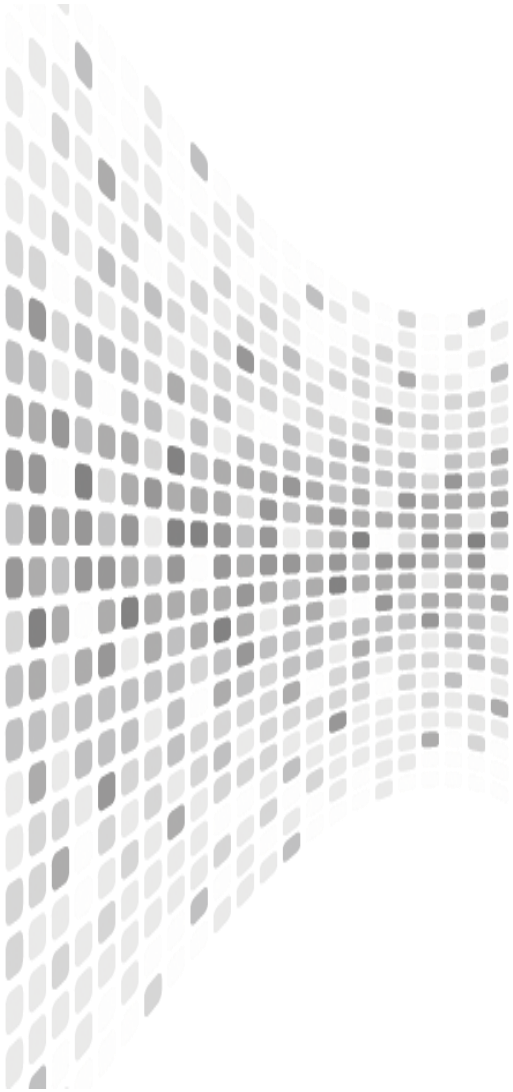


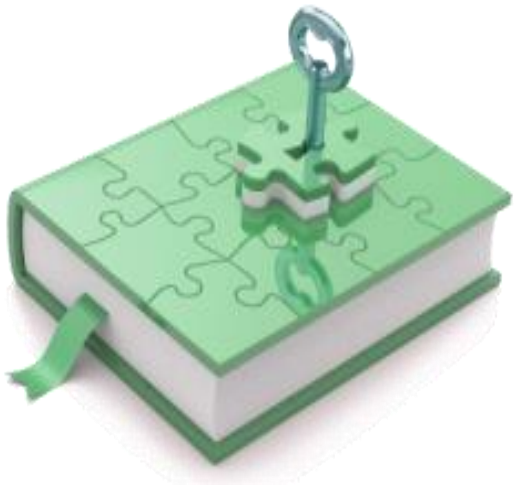
Изображения



Видео







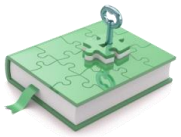
# Данные и Знания



Данные – совокупность зафиксированных фактов



Информация – сведения, уменьшающие неопределённость



Знания – сведения, позволяющие действовать с прогнозируемым результатом

Типичная проблема:

Мы располагаем данными, они хранятся в цифровом виде, но мы не знаем, что в них.

# Что такое машинное обучение?

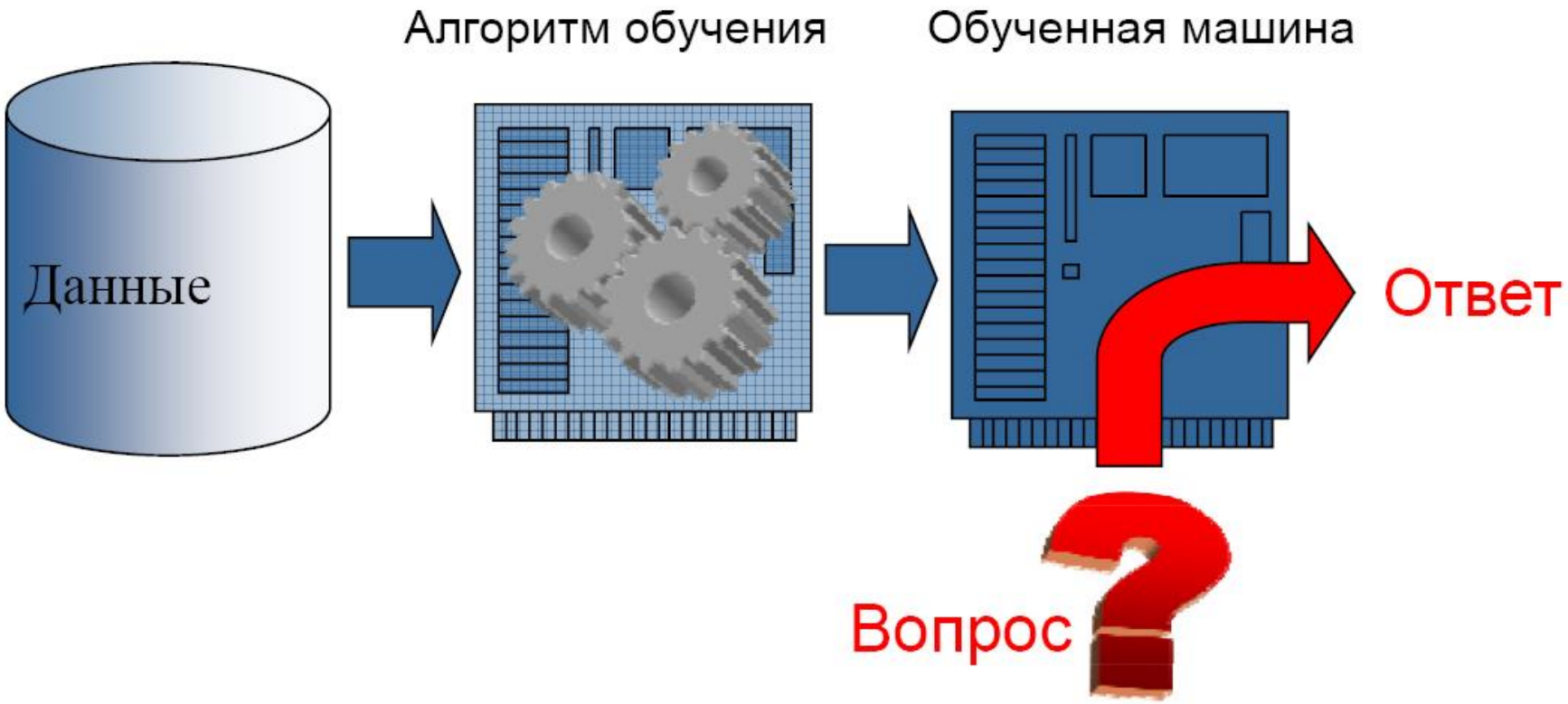
(англ. Machine learning)

подраздел искусственного интеллекта (ИИ), изучающий методы построения **алгоритмов, способных обучаться.**



**Машинное обучение** — это алгоритмы, которые находят в данных скрытые закономерности

# Что такое машинное обучение?



***Машинное обучение — обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться.***

www.machinelearning.ru/wiki/index.php?title=Машинное_обучение

***Машинное обучение — процесс, в результате которого машина (компьютер) способна показывать поведение, которое в нее не было явно заложено (запрограммировано).***

A.L. Samuel Some Studies in Machine Learning Using the Game of Checkers // IBM Journal. July 1959. P. 210-229.

***Говорят, что компьютерная программа обучается на основе опыта E по отношению к некоторому классу задач T и меры качества P, если качество решения задач из T, измеренное на основе P, улучшается с приобретением опыта E.***

T.M. Mitchell Machine Learning. McGraw-Hill, 1997.

# Машинное обучение



# Связь с другими дисциплинами

1. Для изучения машинного обучения надо понимать основы теории вероятности, математической статистики и математического анализа,
2. Также нужно уметь уверенно программировать, желательно - на Python.
3. Неплохо бы владеть хорошим письменным английским.
4. Ни один проект по машинному обучению не обходится без знания предметной области.

# Что нужно для машинного обучения?

Развитие машинного обучения обусловлено тремя факторами; оно испытывает бум именно сейчас - когда сложились все эти три фактора.

1. Первый - развитие математического аппарата и алгоритмов обучения; формулировка многих известных простых задач как задач машинного обучения.
2. Второй - развитие компьютерной техники, так как машинное обучение (а особенно - глубокое) - это очень затратный процесс в вычислительном плане.
3. Третий - сбор большого количества данных; чем больше данных - тем шире применение машинного обучения.

Машинное обучение стоит применять в тех сферах, где традиционные алгоритмы неэффективны.

Традиционные сферы применения - распознавание образов, сложная диагностика, имитация интеллектуальной деятельности, принятие решений.

# Машинное обучение

**Что это**

- область науки на стыке математики и информатики
- изучает алгоритмы, задачи и их свойства
- не имеет отношения к предметной области

**Зачем это**

- автоматизация рутинных процессов
- обработка больших объемов данных
- задачи, которые не могут быть решены явно

# Машинное обучение

- Пример: обнаружения спама

Входные данные:

Множество писем



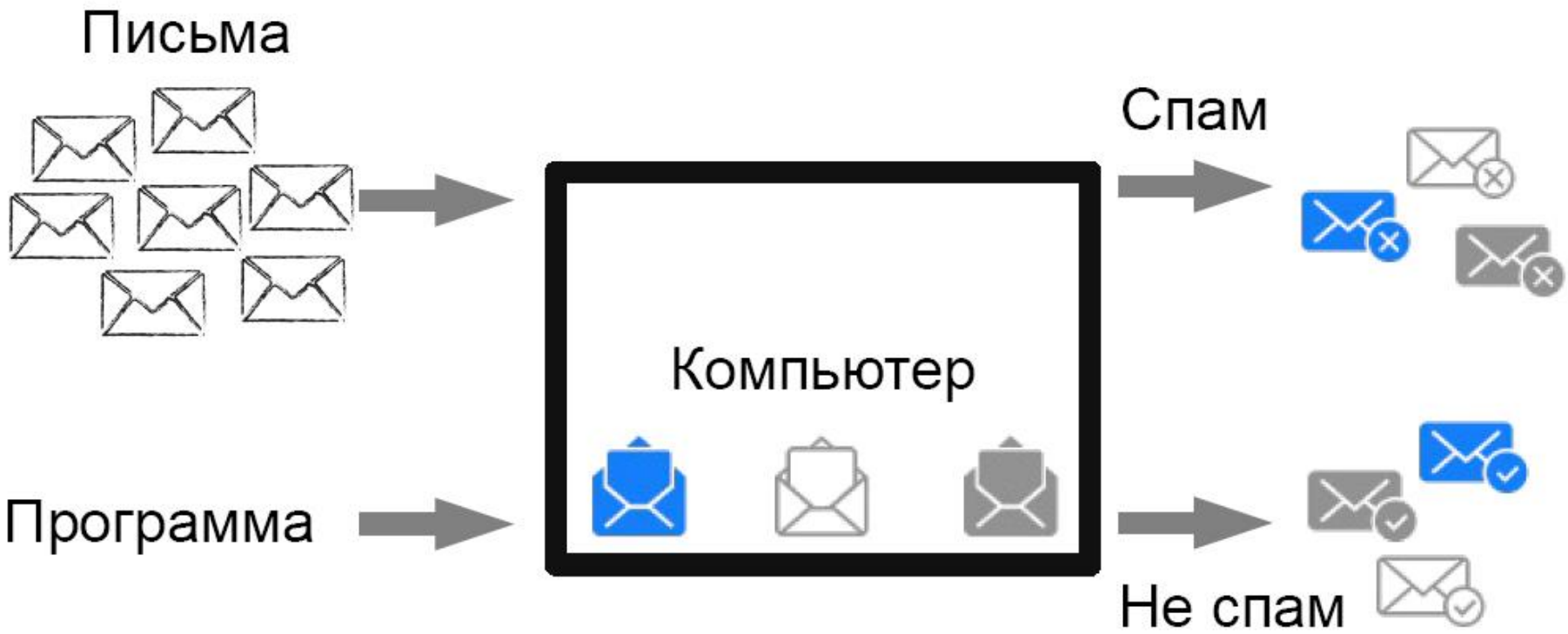

Выходные данные:

Метки для писем из входных данных

(спам/ не спам)

# Машинное обучение

- Пример: обнаружения спама



# Машинное обучение

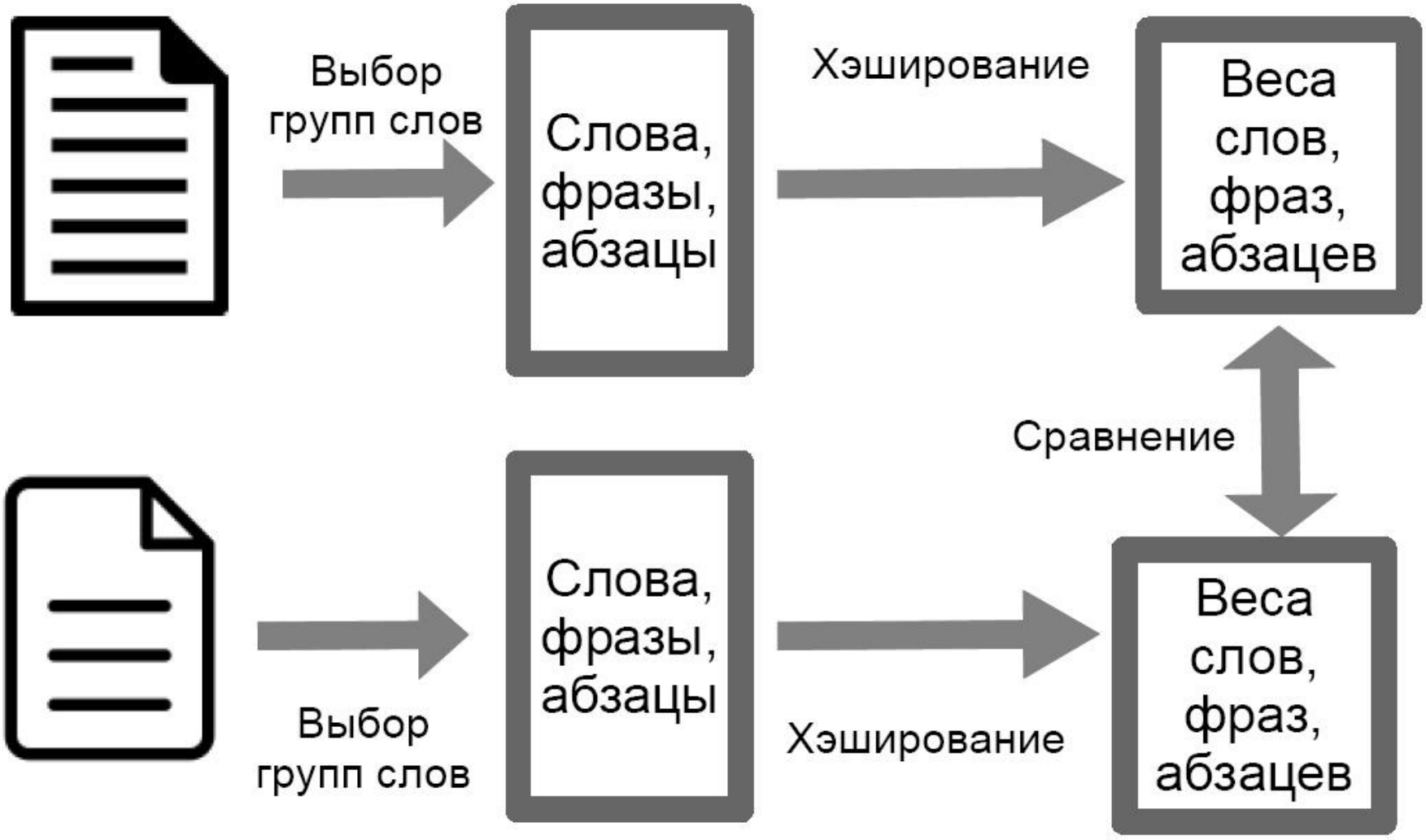
Примеры задач

- Поиск похожих элементов
  - Сайты-двойники
  - Плагиат
- Обработка естественного языка
  - Выделение документов со схожими темами
  - Выделение ключевых слов
- Самонастраивающиеся приложения
  - Рекомендательные системы (Ozon, Netflix)

# Машинное обучение

- Примеры задач

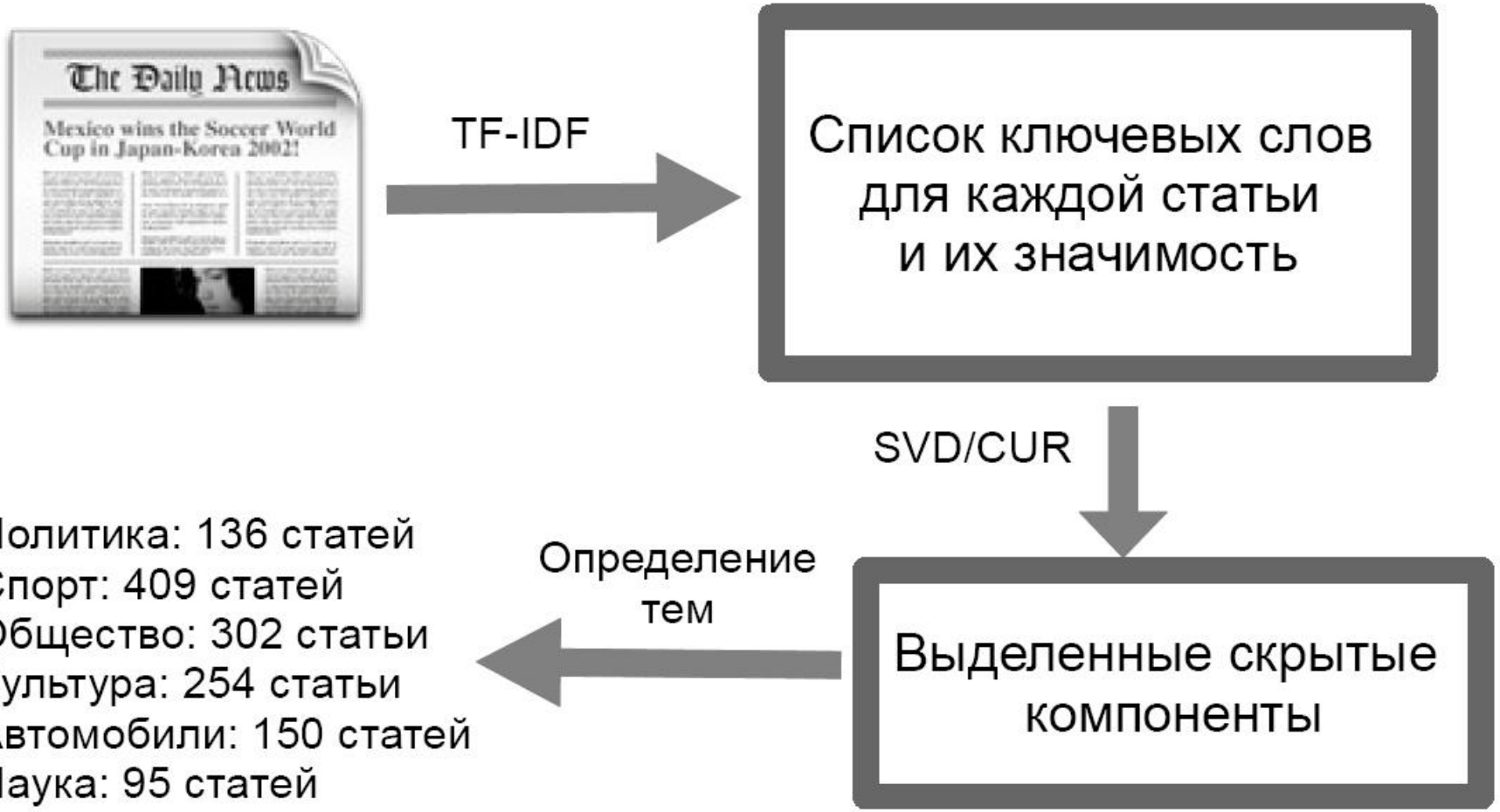
Поиск плагиата



# Машинное обучение

- Примеры задач

Выделение документов со схожими темами



# Вернемся к определению

***Говорят, что компьютерная программа обучается на основе опыта E по отношению к некоторой задаче T и меры качества P, если качество решения задачи T, измеренное на основе P, улучшается с приобретением опыта E.***

- ➔Что такое опыт ***E***?
- ➔ Какие бывают задачи ***T***?
- ➔ Как выбрать критерий качества ***P***?

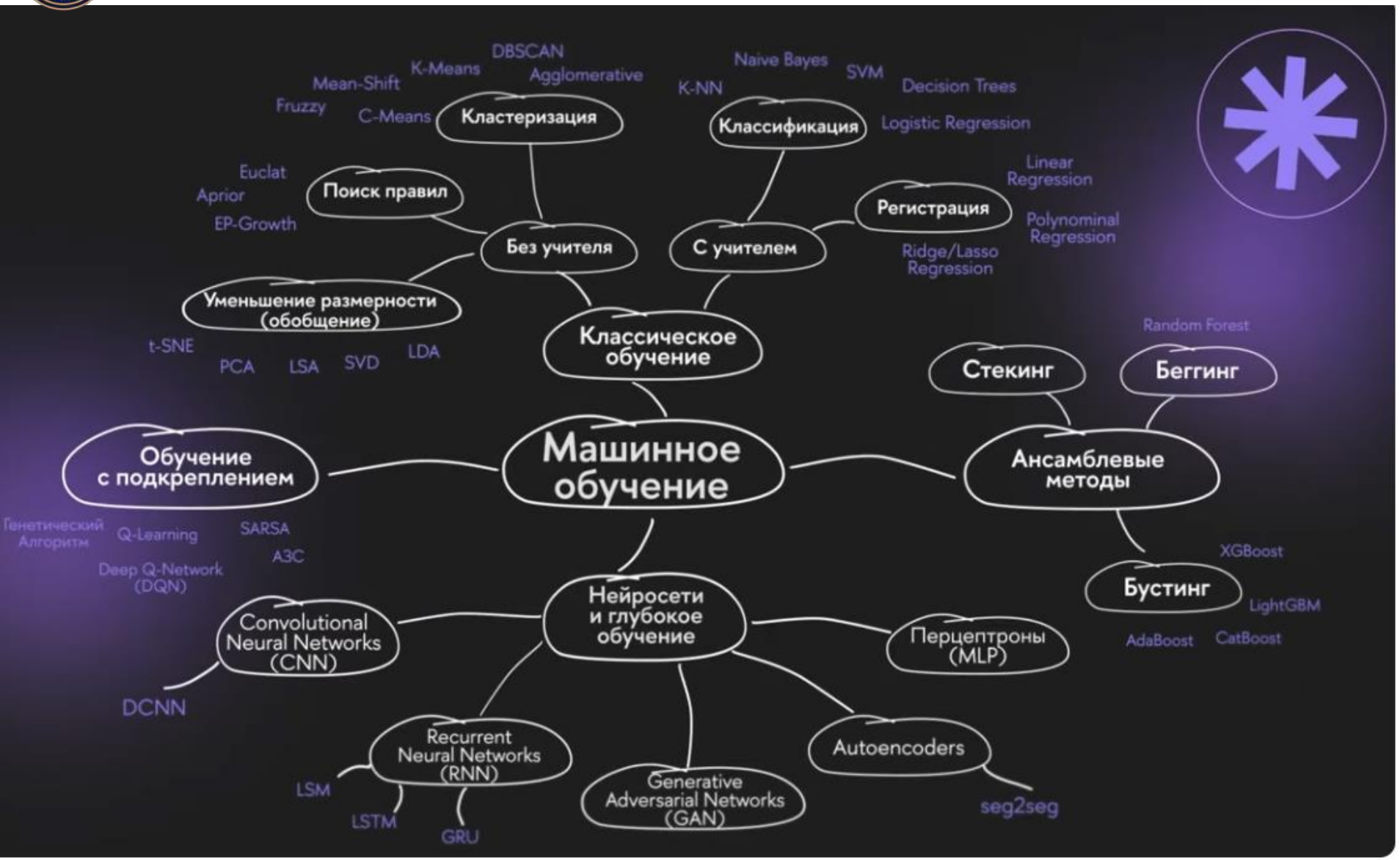
# Пример к определению

## Задача: диагностика опухоли головного мозга по мрт

- ➔ Опыт ***E***: набор пар *<мрт головного мозга, наличие опухоли>*
- ➔ Задача ***T***: построить алгоритм ***A*** идентифицирующий наличие или отсутствие опухоли головного мозга по мрт: ***A****(мрт головного мозга) => наличие опухоли*
- ➔ Критерий качества ***P***: доля правильно поставленных диагнозов

# Что конкретно значит “обучение”?

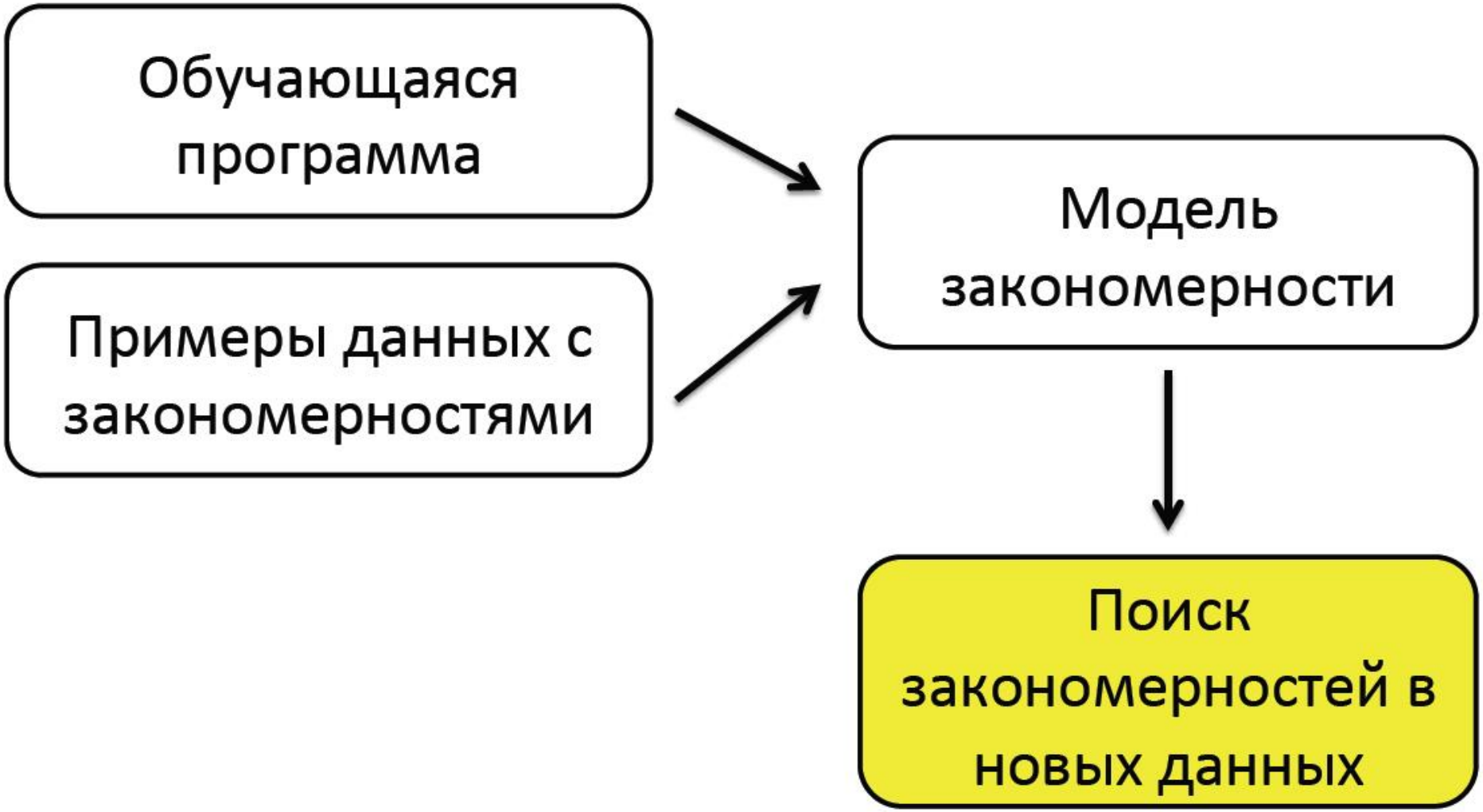
- **Данные** - это как раз и есть **опыт**, на котором модель учится.
- **Модель** - это некоторая **функция** (в широком смысле, не обязательно математическая формула), которая имеет численные параметры.
- **Метрики** - это какой-то способ измерить то, насколько хорошо или плохо данная модель подстроилась под данные (**качество**).


Машинное обучение
Классическое обучение
Кластеризация
K-Means
DBSCAN
Agglomerative
Mean-Shift
Fruzzy
C-Means
Классификация
Naive Bayes
SVM
K-NN
Decision Trees
Logistic Regression
Поиск правил
Euclat
Aprior
EP-Growth
Без учителя
С учителем
Регистрация
Linear Regression
Polynominal Regression
Ridge/Lasso Regression
Уменьшение размерности (обобщение)
t-SNE
PCA
LSA
SVD
LDA
Обучение с подкреплением
Генетический Алгоритм
Q-Learning
SARSA
A3C
Deep Q-Network (DQN)
Ансамблевые методы
Стекинг
Беггинг
Random Forest
Бустинг
XGBoost
LightGBM
AdaBoost
CatBoost
Нейросети и глубокое обучение
Convolutional Neural Networks (CNN)
DCNN
Перцептроны (MLP)
Recurrent Neural Networks (RNN)
LSM
LSTM
GRU
Generative Adversarial Networks (GAN)
Autoencoders
seg2seg

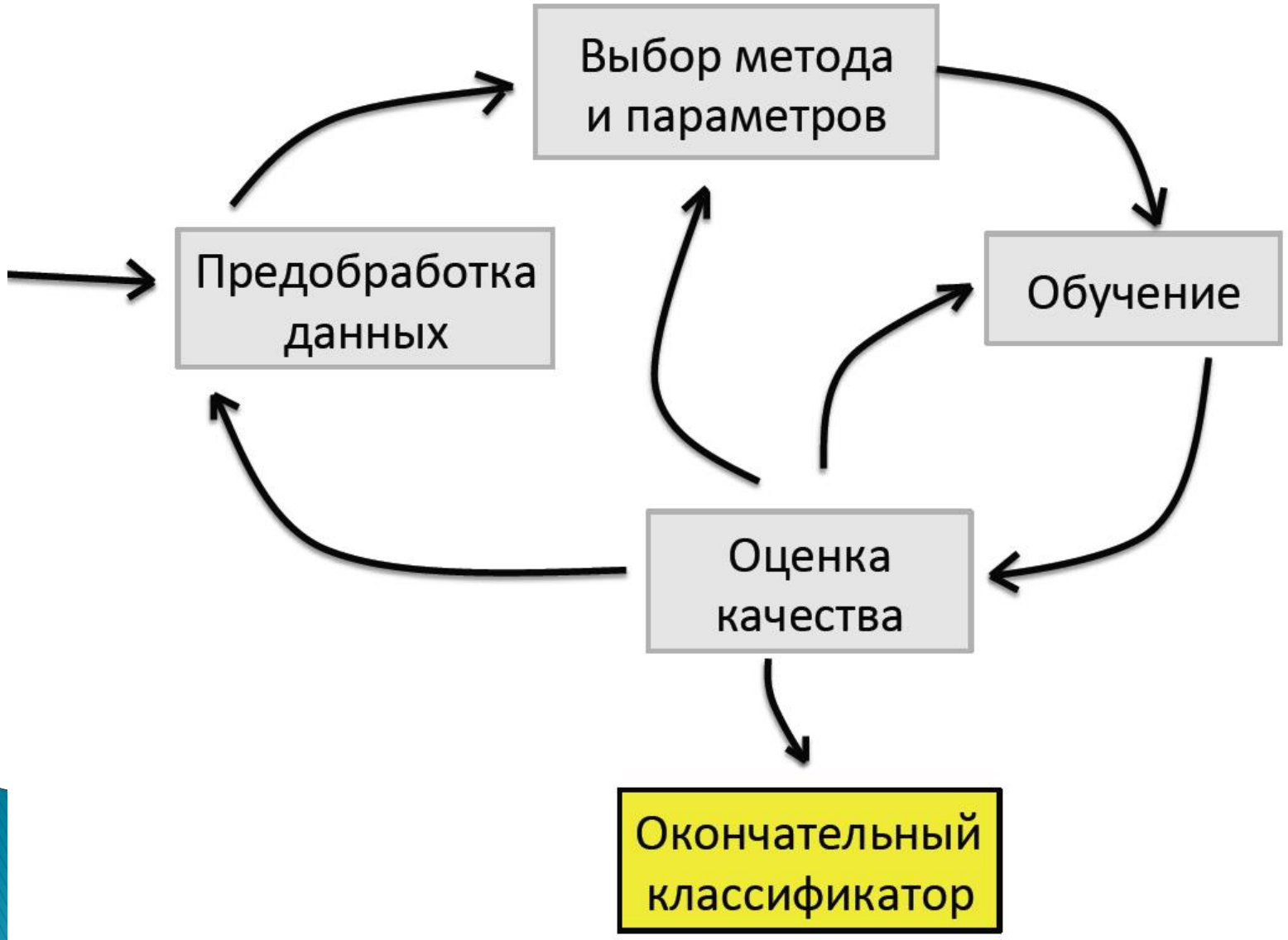
# Способы машинного обучения



# Суть машинного обучения



# Цикл решения задачи

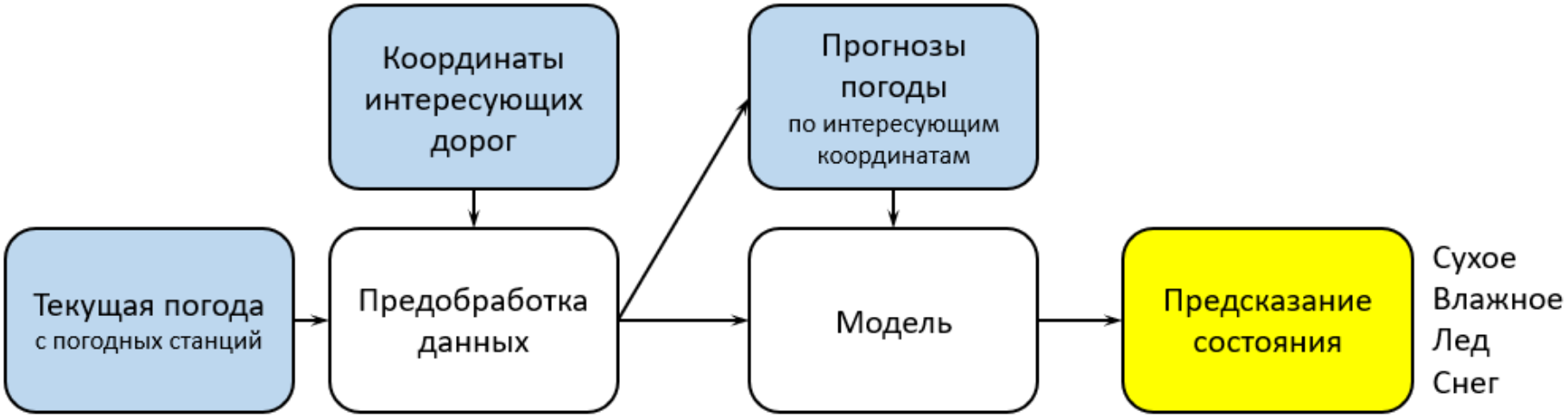


# Что еще может ML?

- Классифицировать
- Давать вещественный ответ (регрессия)
- Прогнозировать
- Ранжировать
- Фильтровать выбросы
- Находить наиболее значимые показатели
- Давать рекомендации

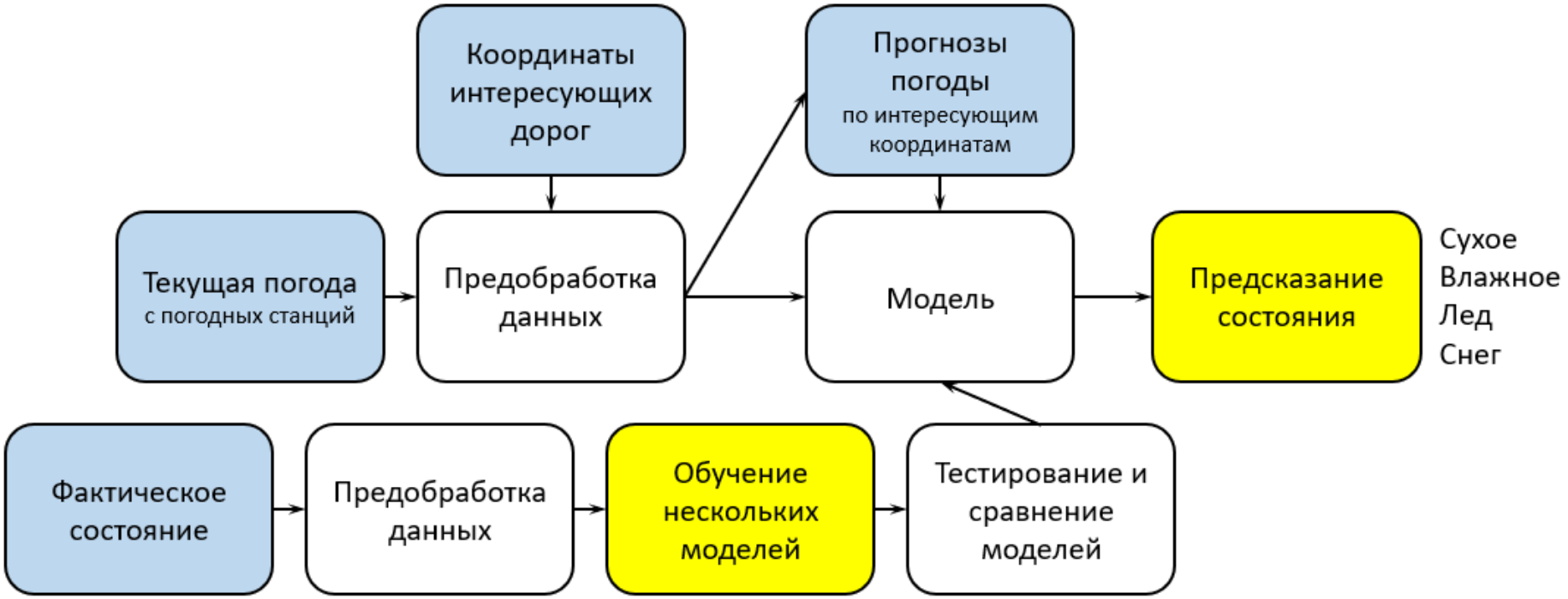
# Пример из жизни

Прогнозирование состояния дорожного полотна на 6 и 24 часа



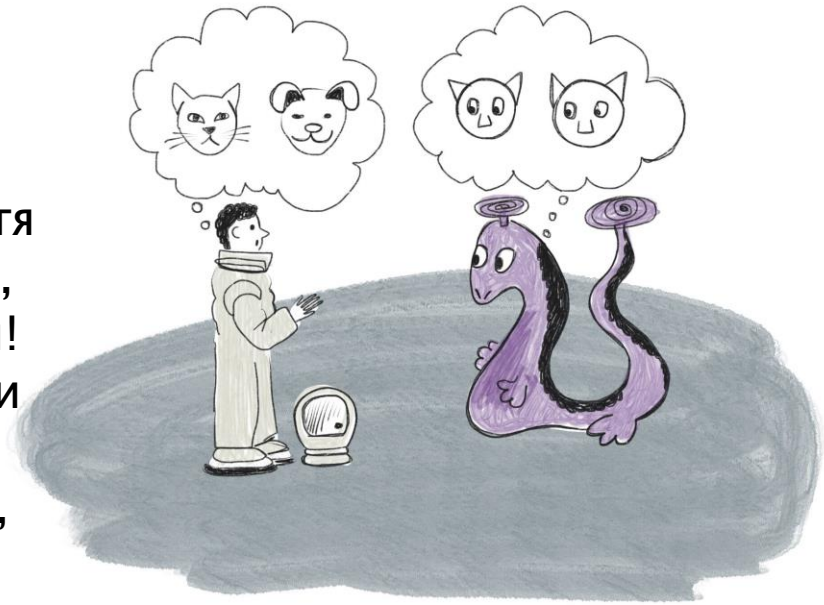
# Пример из жизни

Прогнозирование состояния дорожного полотна на 6 и 24 часа



# Как оличить кошку от собаки?

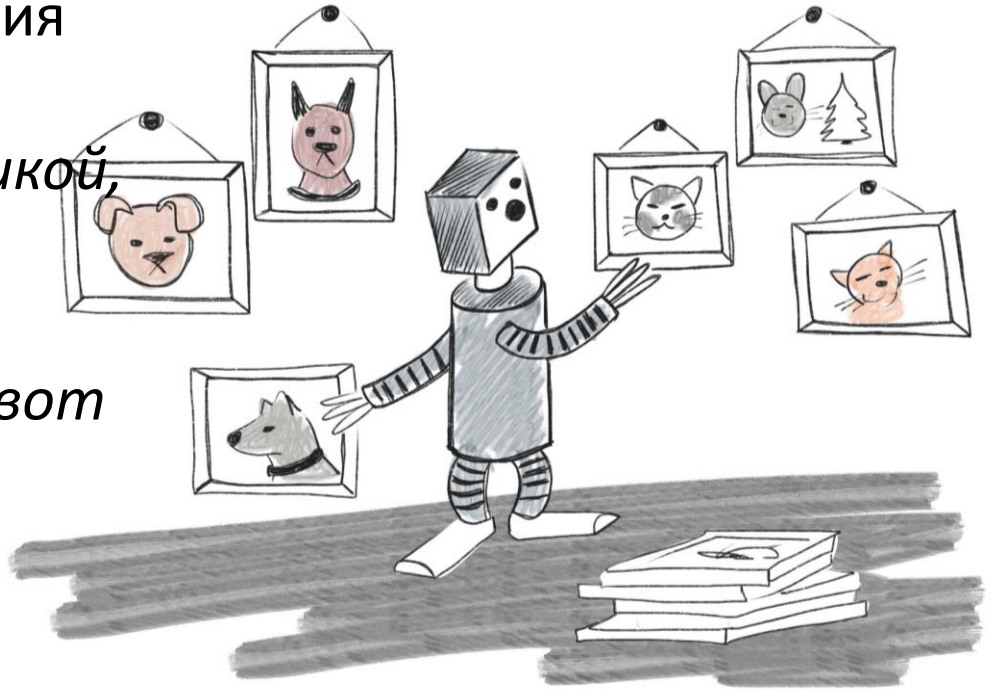
Это юпитерианин Йжун, и он никогда в жизни не видел ни кошек, ни собак, и хочет понять чем они отличаются на вид друг от друга.
– У собаки четыре ноги, – говорите вы.
– А у кошки? – уточняет Йжун.
– Тоже четыре. И ещё есть хвост. Хотя и у собаки есть хвост. Но вот шерсть, шерсть у кошки другая, чем у собаки!
– Она другого цвета? Она короче или длиннее?
– Да нет... – теряетесь вы. – и кошки, и собаки бывают разного цвета и с разной длиной шерсти... И глаз у них по два, и носов по одному...
– Тогда как же вы различаете их?! – удивляется Йжун.



# Как научить компьютер отличать кошек от собак?

Так же, как обучают человека.
Модель машинного обучения использует
- примеры *(картинка с кошкой, картинка с собакой, еще картинка с собакой..)* и
- ответы (или «разметку») *(вот это кошка, а это собака).*

*Такой подход в машинном обучении называется* ***“обучением с учителем”.***



# Задачи машинного обучения

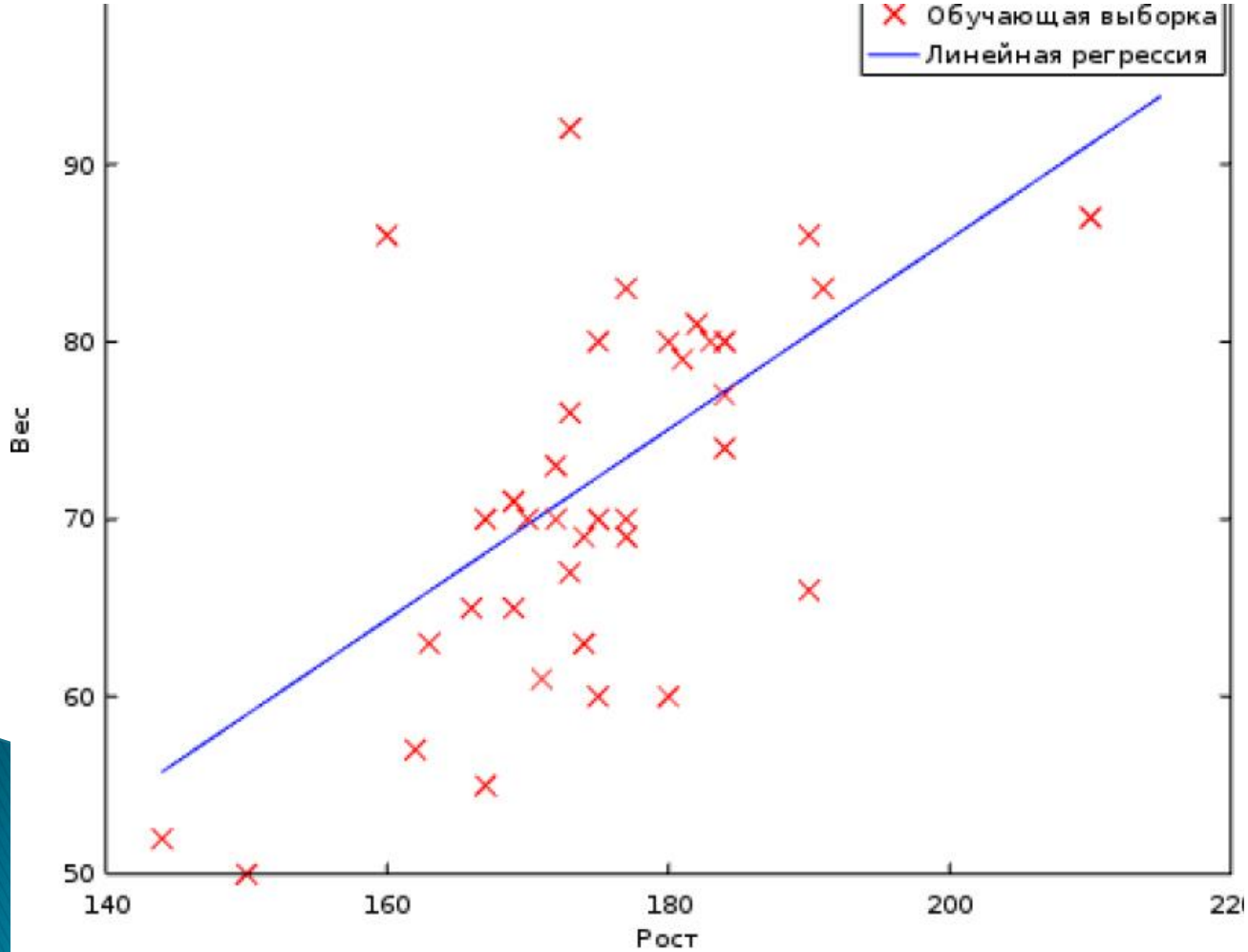
- Обучение с учителем
  - Регрессия
    - Предсказание значения непрерывной величины (например, линейная регрессия)

# Постановка задачи регрессии

1. Регрессия - это задача машинного обучения с учителем, которая заключается в предсказании некоторой непрерывной величины.
2. Для использования регрессионных моделей нужно, чтобы в датасете были характеристики объектов и “правильные” значения целевой переменной.
3. Примеры регрессионных задач - предсказание цены акции, оценка объекта недвижимости.
4. Задача регрессии основывается на предположении, что значение целевой переменной зависит от значения признаков.
5. Регрессионная модель принимает набор значений и выдает предсказание значения целевой переменной.
6. В качестве регрессионных моделей часто берут аналитические функции, например, линейную.

# Задачи машинного обучения

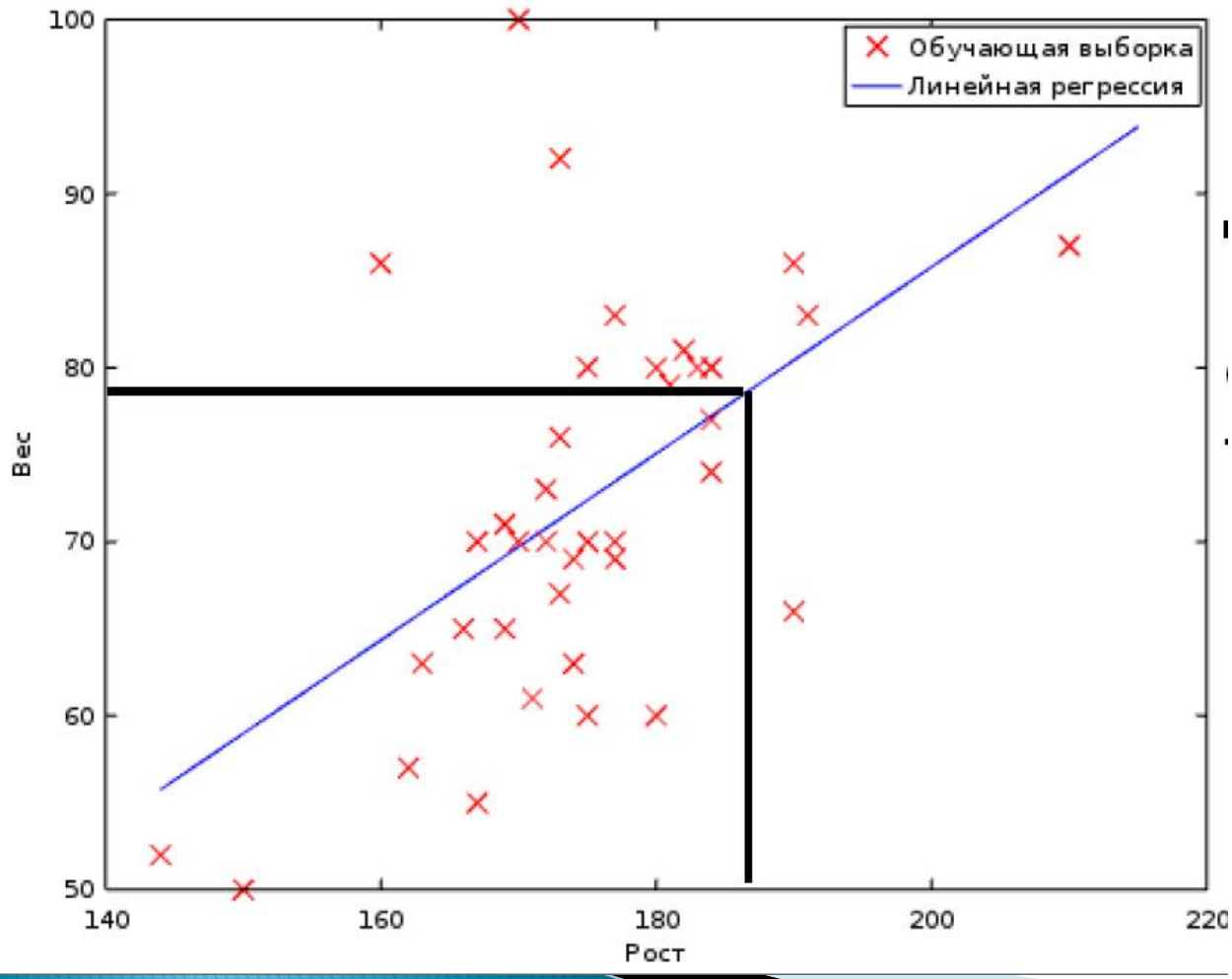
- Обучение с учителем: регрессия




Предсказание веса в зависимости от роста: по обучающей выборке построена линейная регрессия

# Задачи машинного обучения

- Обучение с учителем: регрессия




Для человека с ростом 187 сантиметров теперь можем предсказать вес: 79 килограммов

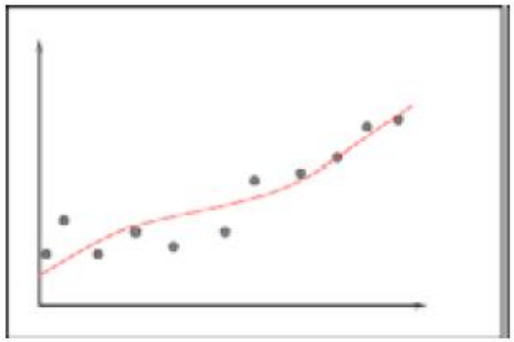
# Задачи машинного обучения

- Обучение с учителем
  - Регрессия
- Предсказание значения непрерывной величины (например, линейная регрессия)
  - Классификация
- Предсказание значений дискретной величины (например, линейная классификация)

# Примеры задач: регрессия

- **Требуется определить возраст моллюска:**
- Можно посмотреть в микроскоп и посчитать число колец раковины
  - Проблема ! Это сложно, долго, дорого
- Идея! Можно посчитать простые признаки: пол, диаметр, высоту, вес раковины, … и предсказать по ним возраст
- **Задача регрессии:**
  - примеры для обучения = (простые измерения, возраст)
  - находим, как зависит возраст от измерений (получаем формулу)
  - для нового примера (возраст не известен) вычисляем значение возраста по формуле





# Обучение без учителя

**Обучение без учителя** – способ машинного обучения, восстанавливающий внутренние взаимосвязи в исходных неразмеченных данных

**Типы входных данных:**
- признаковое описание объектов
- матрица расстояний между объектами

**Типы задач:**
- кластеризация
- понижение размерности
- визуализация данных
- поиск правил ассоциации
- заполнение пропущенных значений

# Задачи обучения без учителя

1. Обучение без учителя - это набор задач по построению внутренней структуры данных, при этом не выделяется целевая переменная.
2. Классические задачи обучения без учителя: кластеризация, уменьшение размерности, обнаружение аномалий, поиск ассоциаций.
3. Существует глубокое обучение без учителя.
4. При обучении без учителя необходимо следить за релевантностью обучающей выборки.
5. При обучении без выбора учителя также можно разделить обучающую и тестовую.
6. Модели без обучения, используемые для извлечения полезных ископаемых из данных.
7. Модели обучения без учителя могут быть вспомогательными при подготовке данных для обучения с учителем.

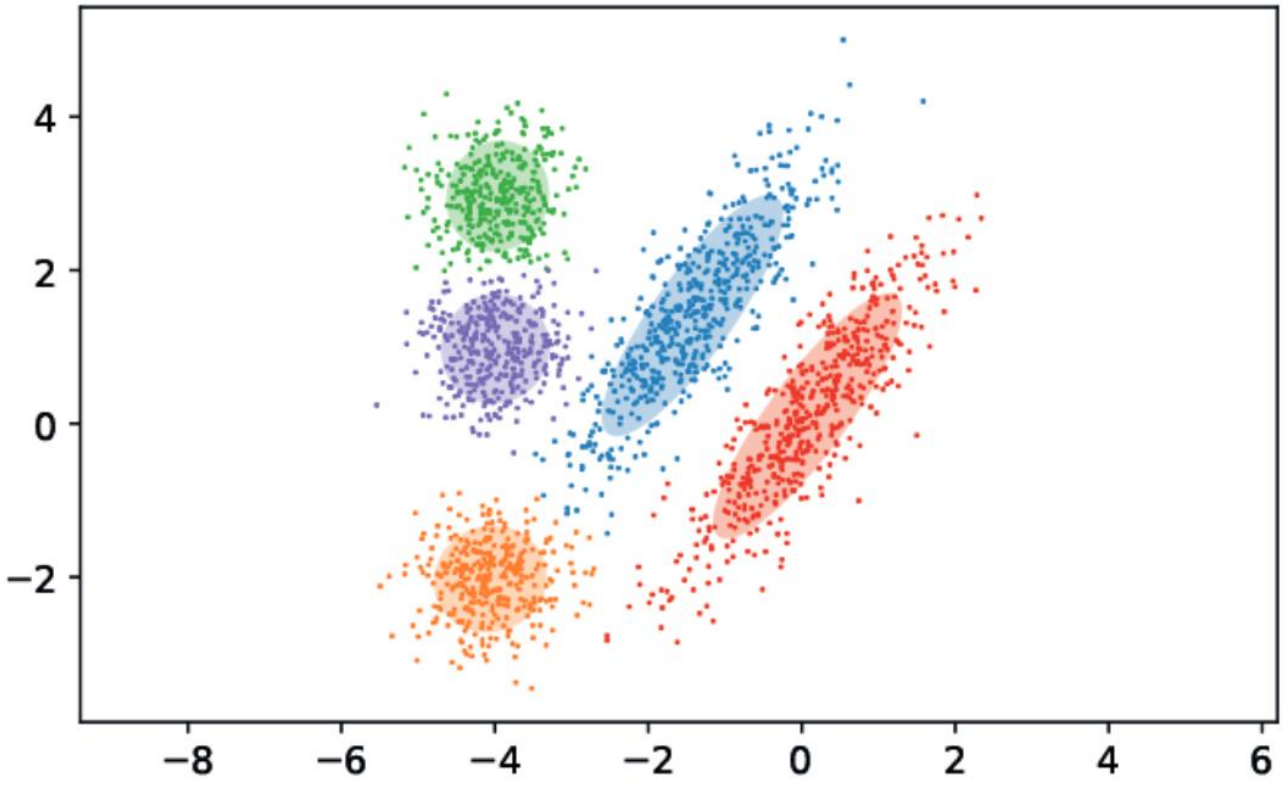
# Кластеризация

1. Кластеризация - разбиение объектов на непересекающиеся кластеры.
2. Кластеров может быть получено различное количество, в некоторых моделях его можно задать заранее.
3. Сами кластеры не несут предметного смысла, его надо искать вручную.
4. Предпочтительно кластеры должны быть однородными.
5. Кластеризация применяется при сегментации клиентов по параметрам.
6. На входе может подаваться матрица признаков объектов либо матрица расстояний между объектами, либо матрица сходства.
7. Метрики дальности или сходства могут применяться по-разному в зависимости от задачи.
8. Для сжатия данных может применяться кластеризация — можно оставить один образ представителю от каждого кластера.

# Кластеризация

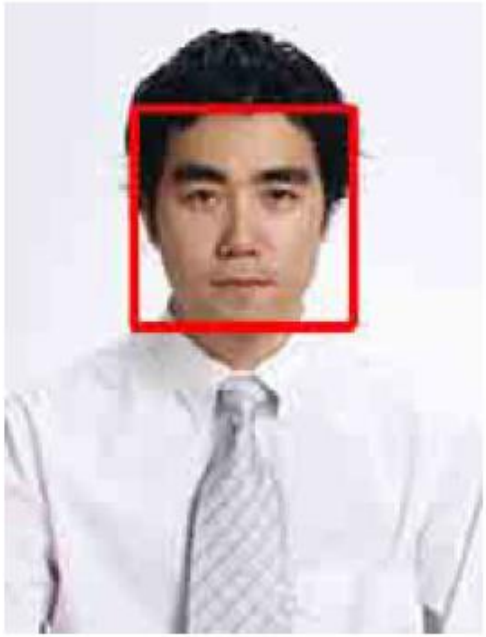
**Алгоритмы кластеризации:**

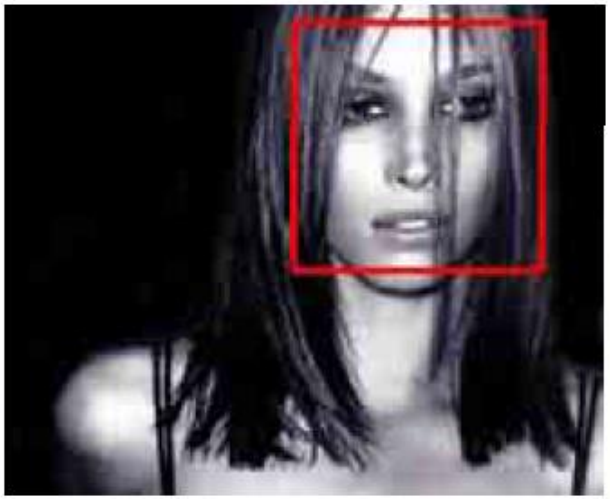
- метод К-средних
- DBSCAN
- иерархическая кластеризация
- …



# Машинное обучение в зрении

- Зрение высокого уровня
  - Обнаружение объектов на изображениях







# Машинное обучение в зрении

- Зрение высокого уровня
  - Распознавание лиц/подписи/отпечатков пальцев/ сетчатки глаза



→ Иван
Иванов

# Машинное обучение в зрении

- Зрение высокого уровня
  - Распознавание печатных/ рукописных символов

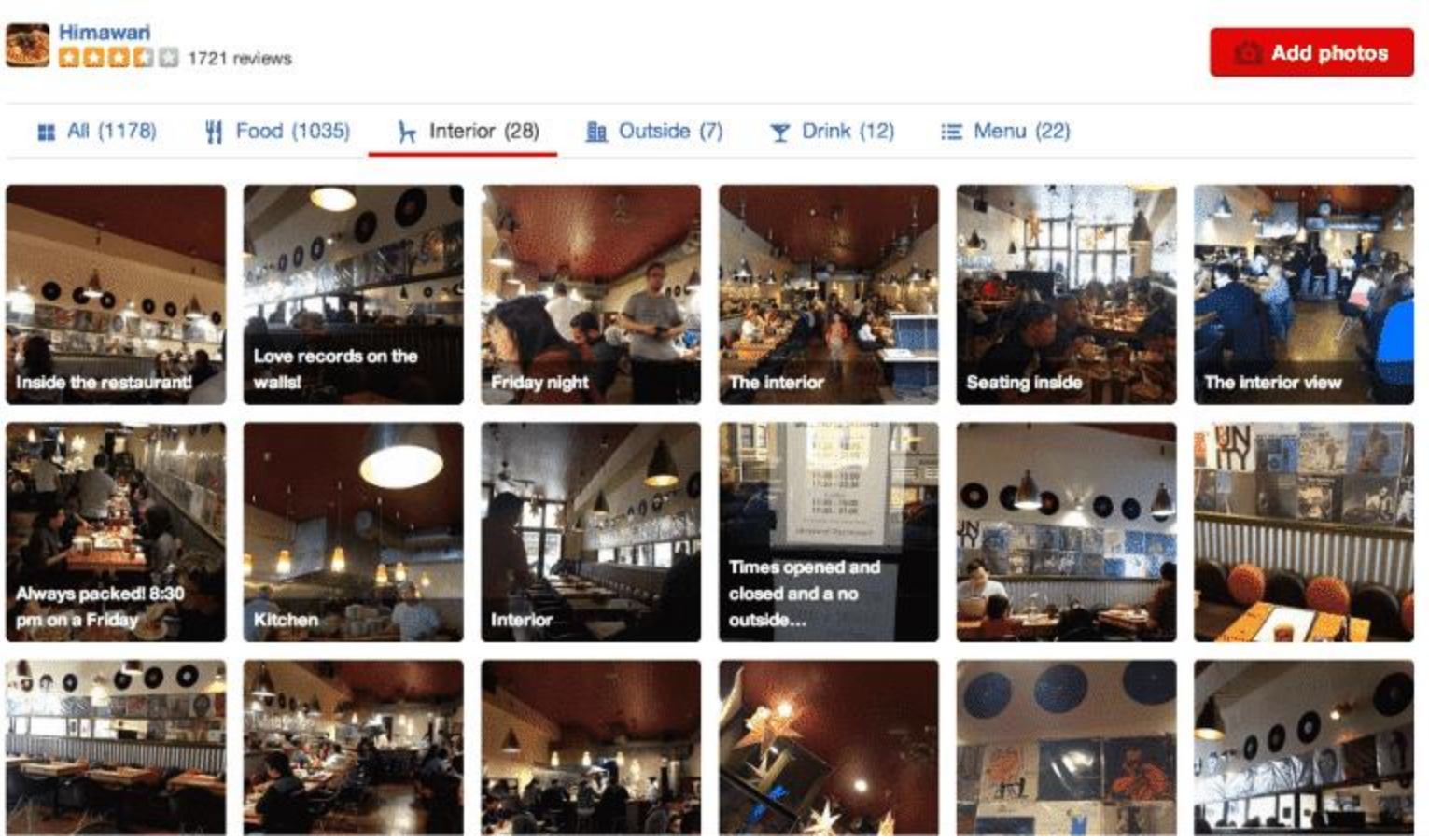


→

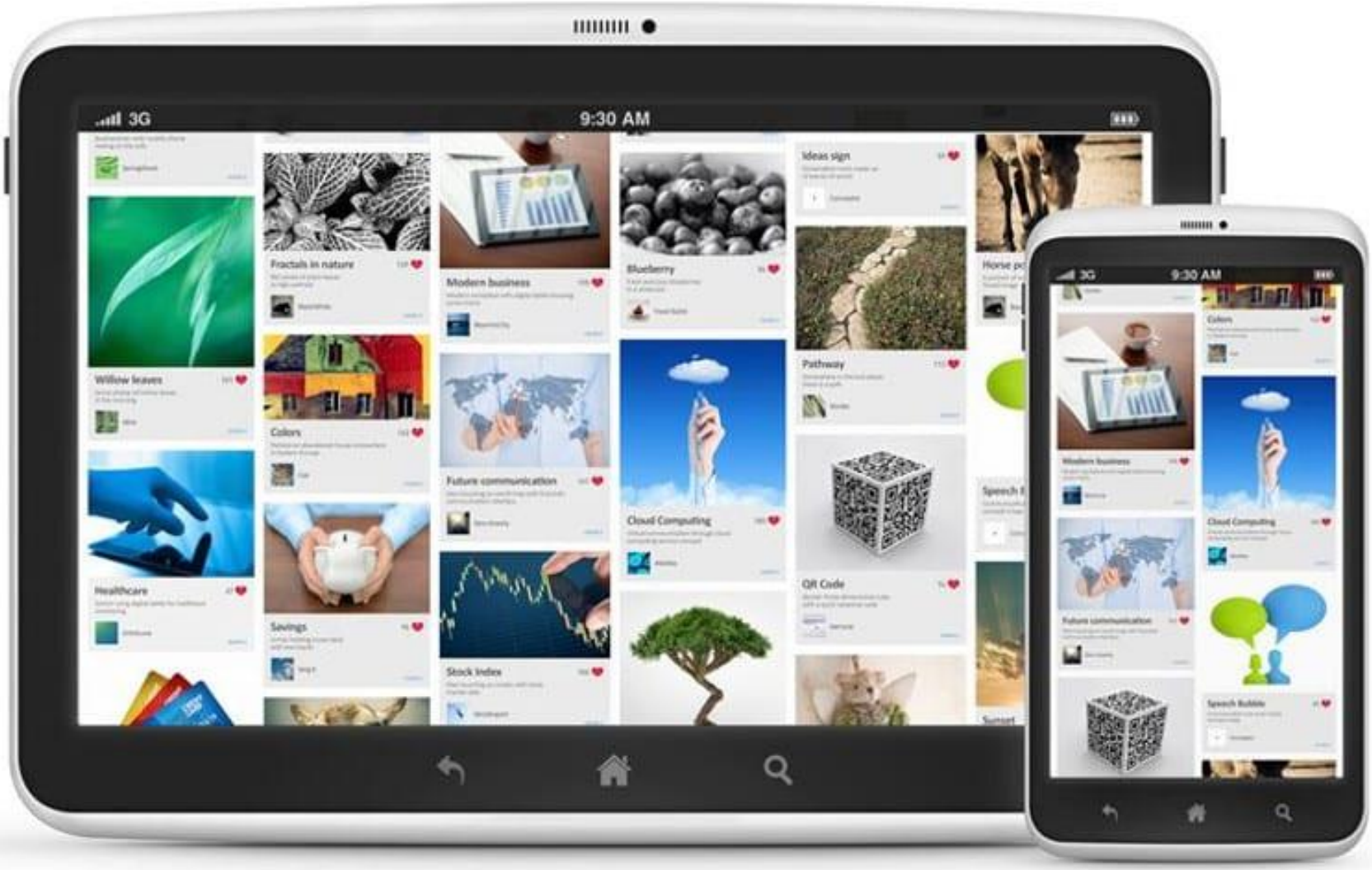
7210414959

…

# Yelp - Курирование изображений



# Pinterest - Поиск контента



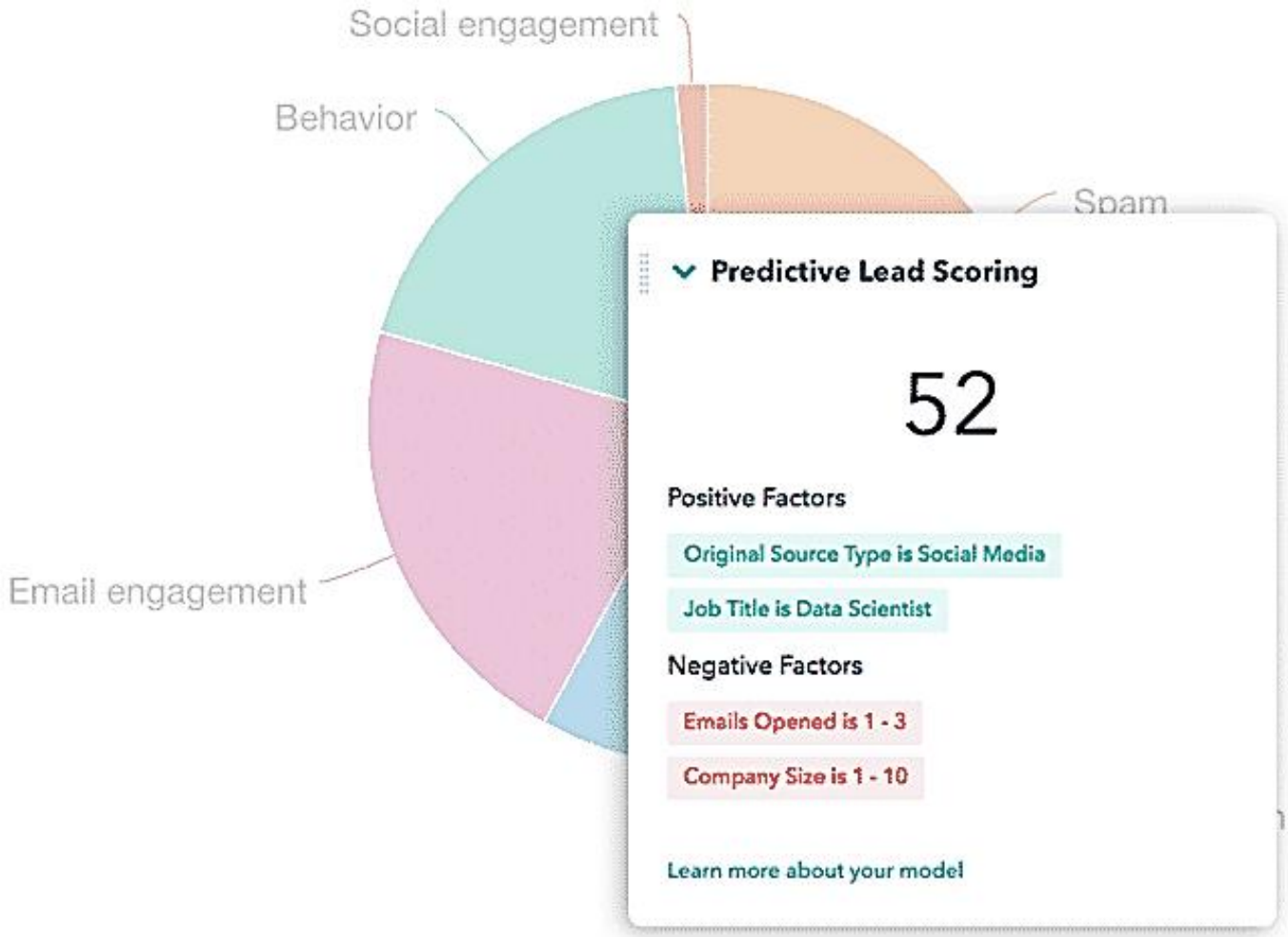
# Facebook - Армия чатботов



# Google - Нейронные сети



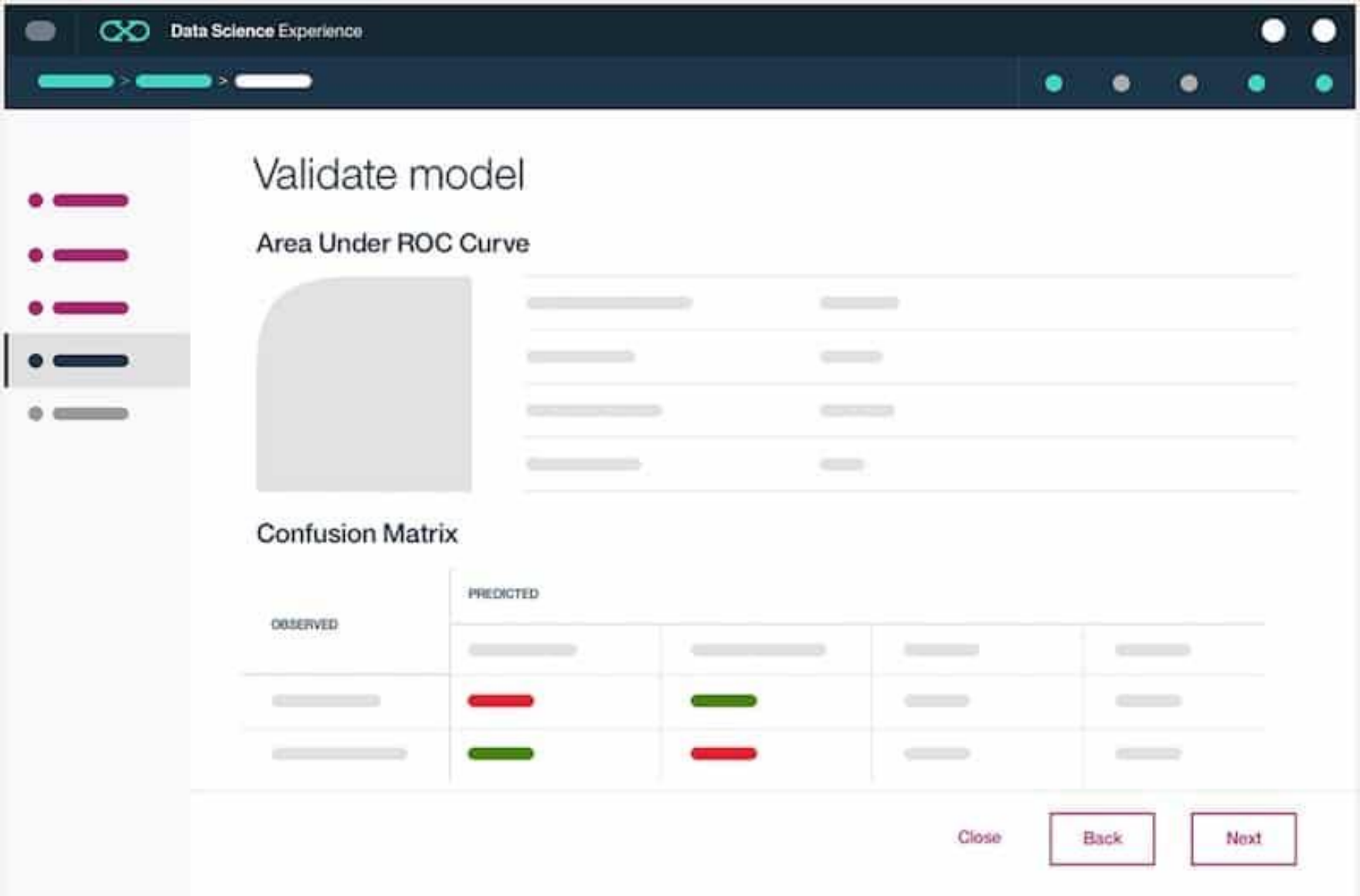
# HubSpot - Умные продажи



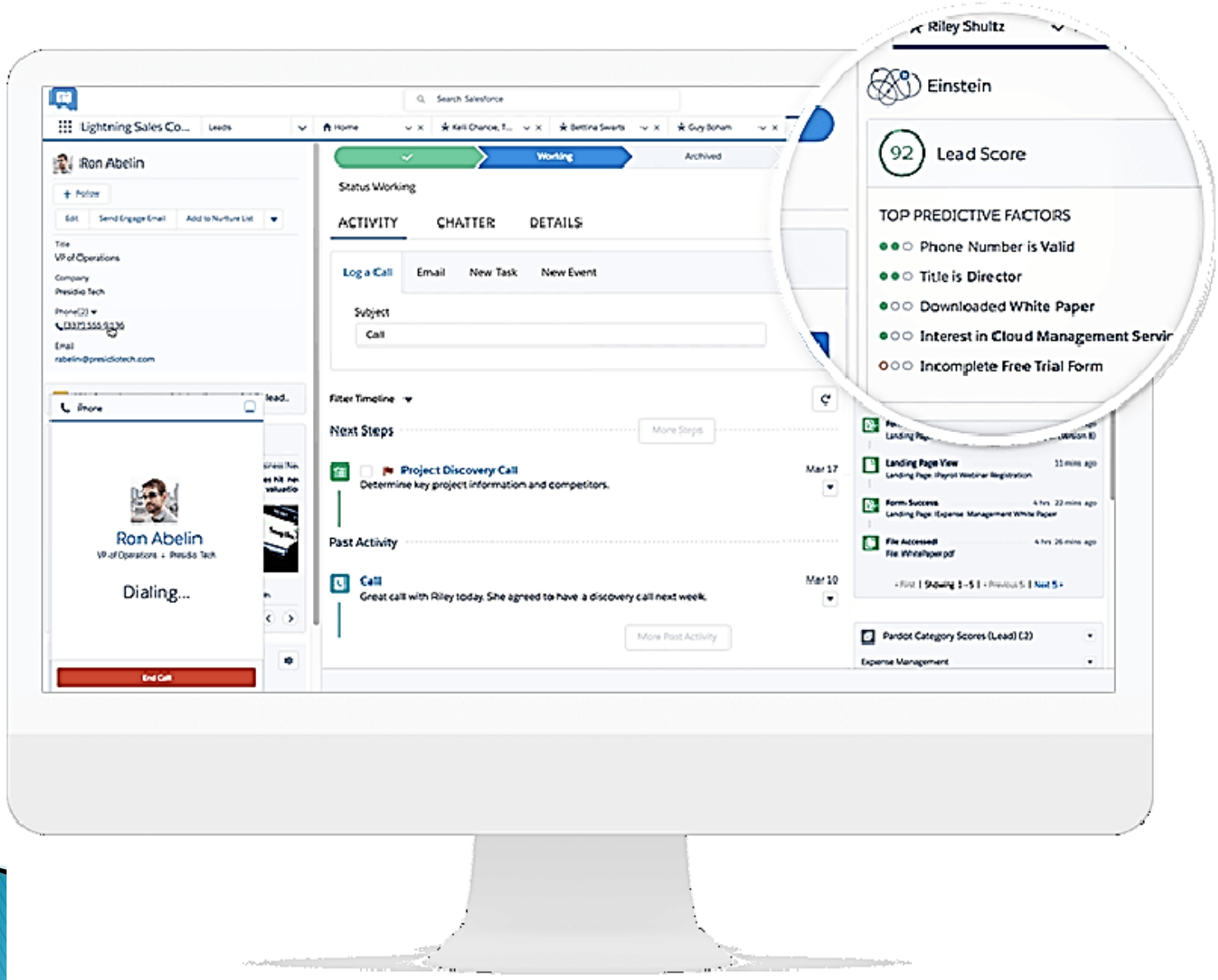

Most influential attributes

Phone Number is valid

# IBM - Здравоохранение нового поколения

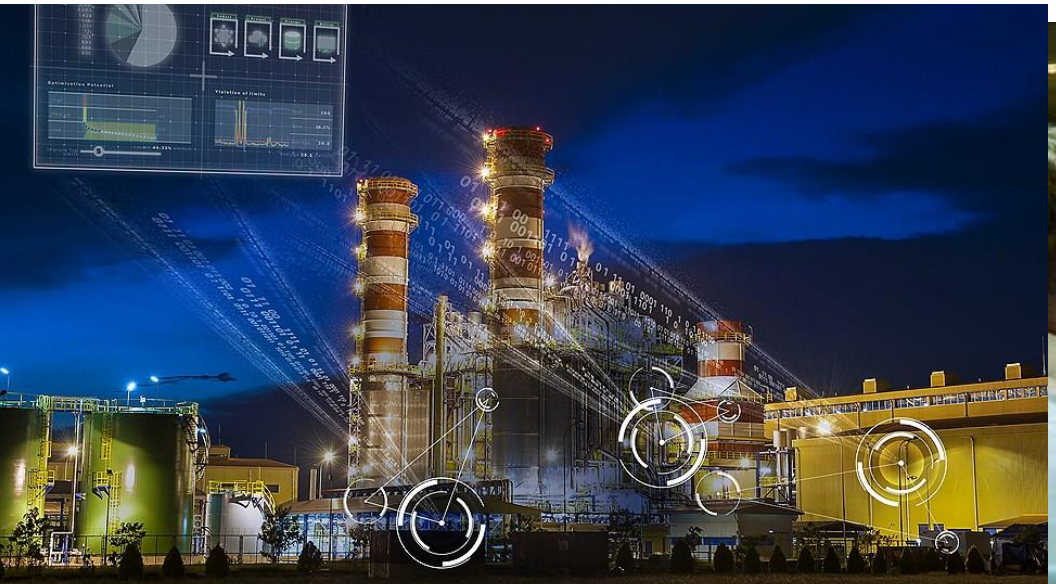


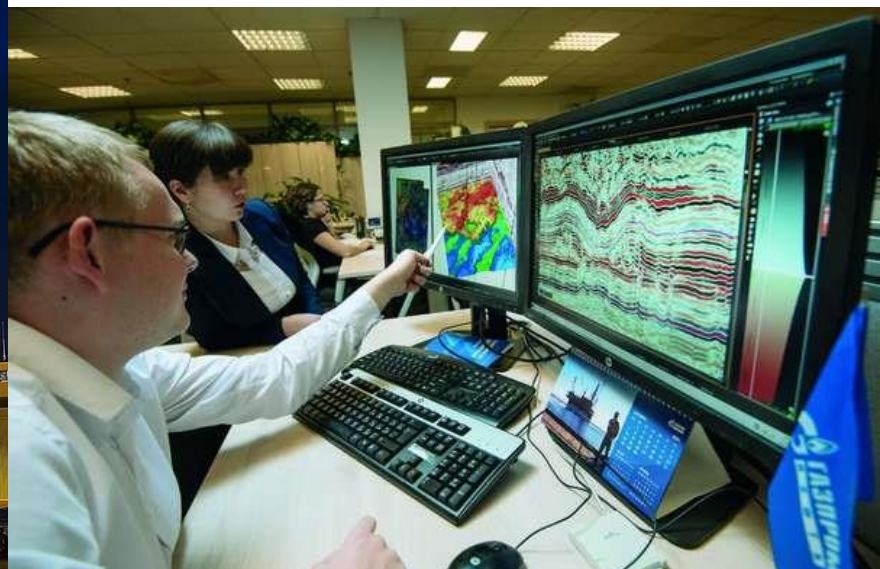
# Salesforce - Умные CRM системы



# Машинное обучение в промышленности

- Минимизация простоев на производстве
- Создание системы управления производством
- Выявление угроз безопасности
- Разведка новых месторождений





# Машинное обучение в финансах

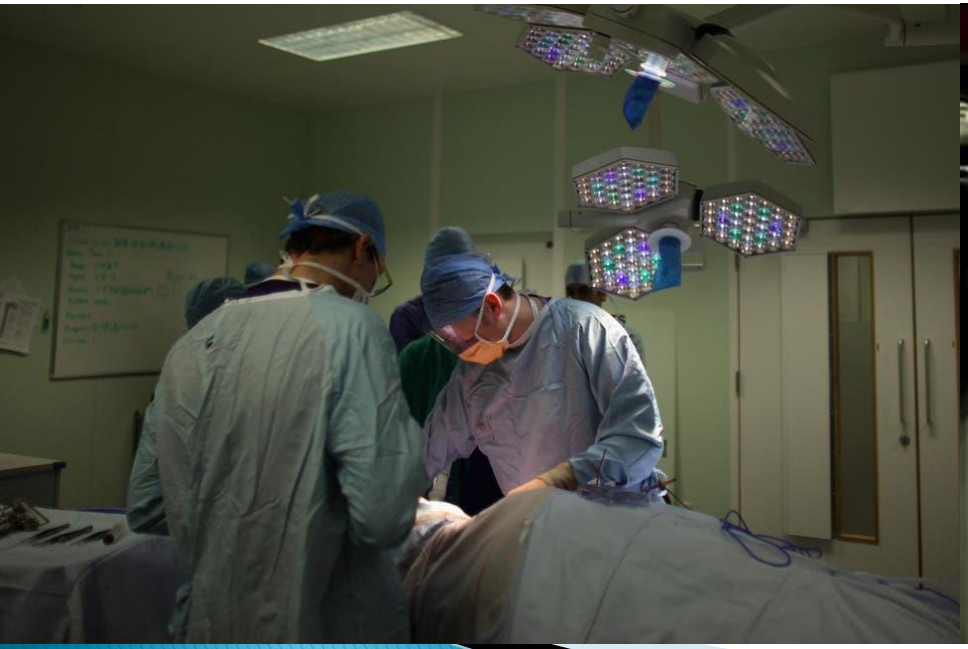
- Оценка кредитоспособности
- Борьба с мошенничеством






# Примеры машинного обучения в медицине

- Улучшение клиентского сервиса
- Диагностика заболеваний
- Автоматические роботизированные операции





# Применение машинного обучения в ритейле и маркетинге

- Прогнозирование действий покупателей, персонализированные предложения и реклама
- Прогнозирование спроса и автоматизация закупок





# Машинное обучение в логистике

- Экономия топлива и повышение производительности транспорта
- Предотвращение сбоев в поставках





# Перспективы применения машинного обучения

- В 2023 году 43% компаний в Европе, США и Китае используют искусственный интеллект и машинное обучение. По оценкам экспертов, к 2025 году рынок машинного обучения вырастет на 42%.



# Зачем следует использовать машинное обучение?

- Christian Heilmann высказал правильную мысль о машинном обучении
- “Можно надеяться, что остальные будут использовать эту мощь только в мирных целях. Я, например, не рассчитываю на эту милость. Предпочитаю играть и быть частью этой революции. И вы присоединяйтесь”.

# БИОМЕТРИЧЕСКИЕ СИСТЕМЫ







## ПРОШЛОЕ

- Обнаружение лиц с помощью группы методов Виолы-Джонса
- Распознавание лиц с использованием обученных признаков
- Обнаружение и распознавание лиц в сложных условиях

## НАСТОЯЩЕЕ

- Детектирование лиц продвинутыми методами предыдущего поколения
- Детектирование лиц в сложных условиях с помощью нейросетей (не real-time)
- Распознавание лиц по изображениям высокого качества с очень высокими вероятностями
- Распознавание лиц в сложных условиях с достаточно высокими вероятностями

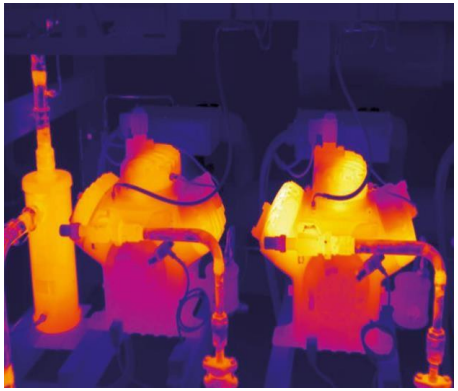
## БУДУЩЕЕ

- Повсеместная замена электронных ключей на распознавание лиц
- Высокие точности определения пола, возраста, национальности, эмоций в задачах маркетинга
- Детектирование и распознавание людей в толпе по большим базам с высокой вероятностью
  -единая биометрическая глобальная система в рамках городов
- Коммерческое внедрение оплаты по лицу

# ПРОМЫШЛЕННАЯ БЕЗОПАСНОСТЬ И АВТОМАТИЗАЦИЯ







| ПРОШЛОЕ | НАСТОЯЩЕЕ | БУДУЩЕЕ |
|---|---|---|
| Обнаружение движения или движущихся объектов в запрещенных зонах (sterile zone) | Распознавание средств индивидуальной защиты (чувствительно к марке) | Автоматические системы контроля за производством |
| Контроль температуры установок с помощью тепловизионных камер | Задачи технологического контроля: определение качества продукции, правильности течения технологических процессов и пр. | Автоматические интеллектуальные системы управления заводами: анализ показаний датчиков системы, предсказание чрезвычайных ситуаций, принятие решений |
| Замена человека в простейших действиях автоматическими системами (задачи подсчета продукции, считывания маркировок/штрихкодов и пр.) | Поиск полезных ископаемых по большим данным – набору показаний группы датчиков | Создание автономных подвижных роботов нового поколения с системой технического зрения на борту |
| | Автоматизированные заводы с минимальным количеством персонала | |

# АНАЛИЗ БОЛЬШИХ ДАННЫХ



**Объем данных, собранных только с двигателей коммерческих реактивных самолетов США в течение года**

**= 1 041 600 500 ТВ**

**НАСТОЯЩЕЕ - БУДУЩЕЕ**

- Анализ состояния бортового авиационного оборудования в масштабе реального времени с возможностью прогнозирования различных типов отказов.
- Совершенствование средств прогнозирования времени прибытия воздушных судов в составе систем управления воздушным движением.




**НАСТОЯЩЕЕ - БУДУЩЕЕ**

- Глубокие нейронные сети в инвестировании
- Предсказание эффективности компаний в терминах выручки, операционных доходов и чистой прибыли на основе финансовой информации и патентах компаний
- Предсказание поведения фондового рынка

***и еще множество приложений…***

# Спасибо за внимание!