

Машинное обучение

Лекция 1

Введение в машинное обучение

Михаил Гущин

mhushchyn@hse.ru

НИУ ВШЭ, 2023



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Организационные моменты



Наша команда

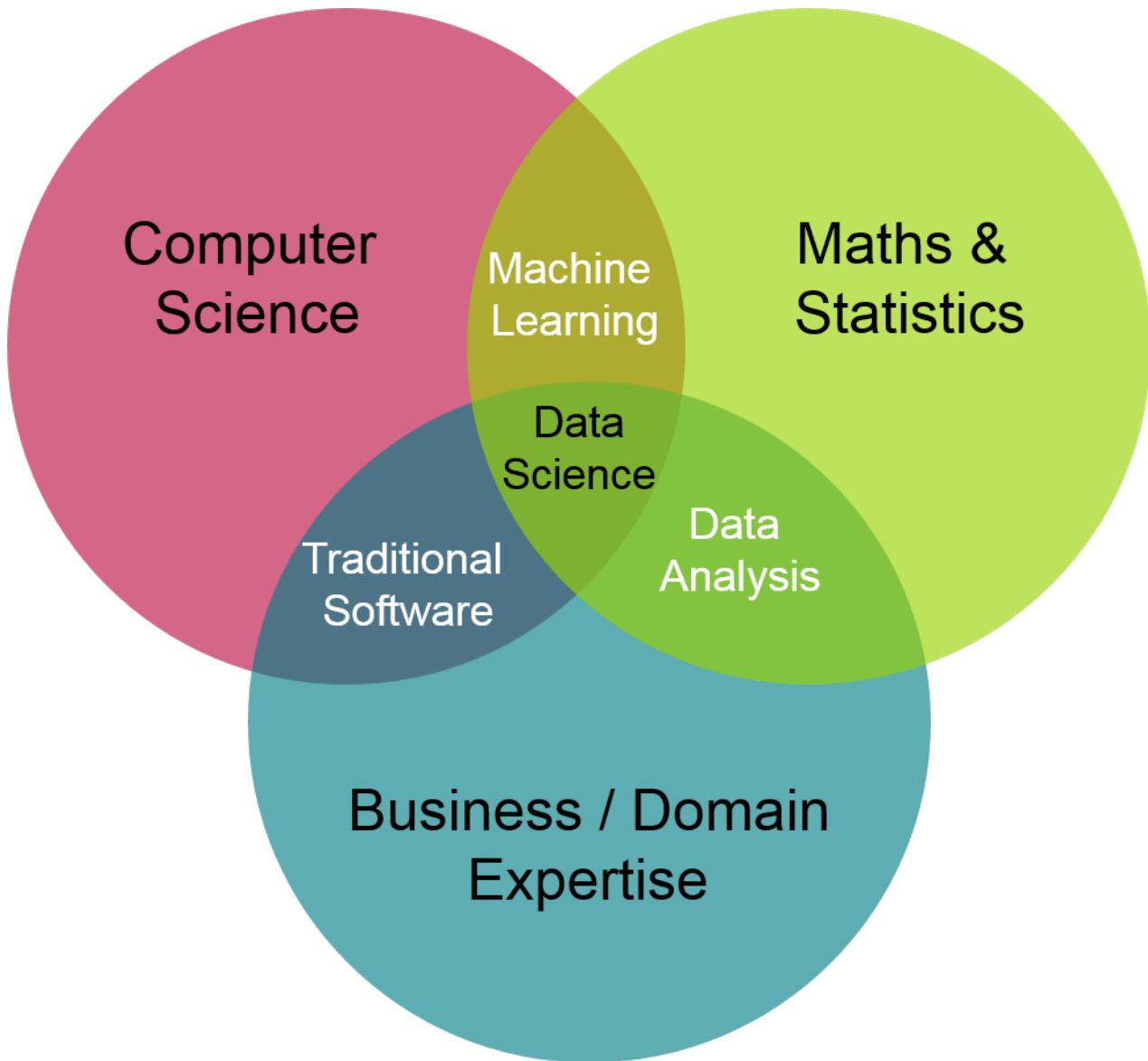
- ▶ **Лектор:**
 - Гущин Михаил
- ▶ **Семинаристы:**
 - Пономарев Алексей
 - Гущин Михаил
- ▶ **Ассистенты:**
 - Сафонов Кирилл
 - Фурик Варвара

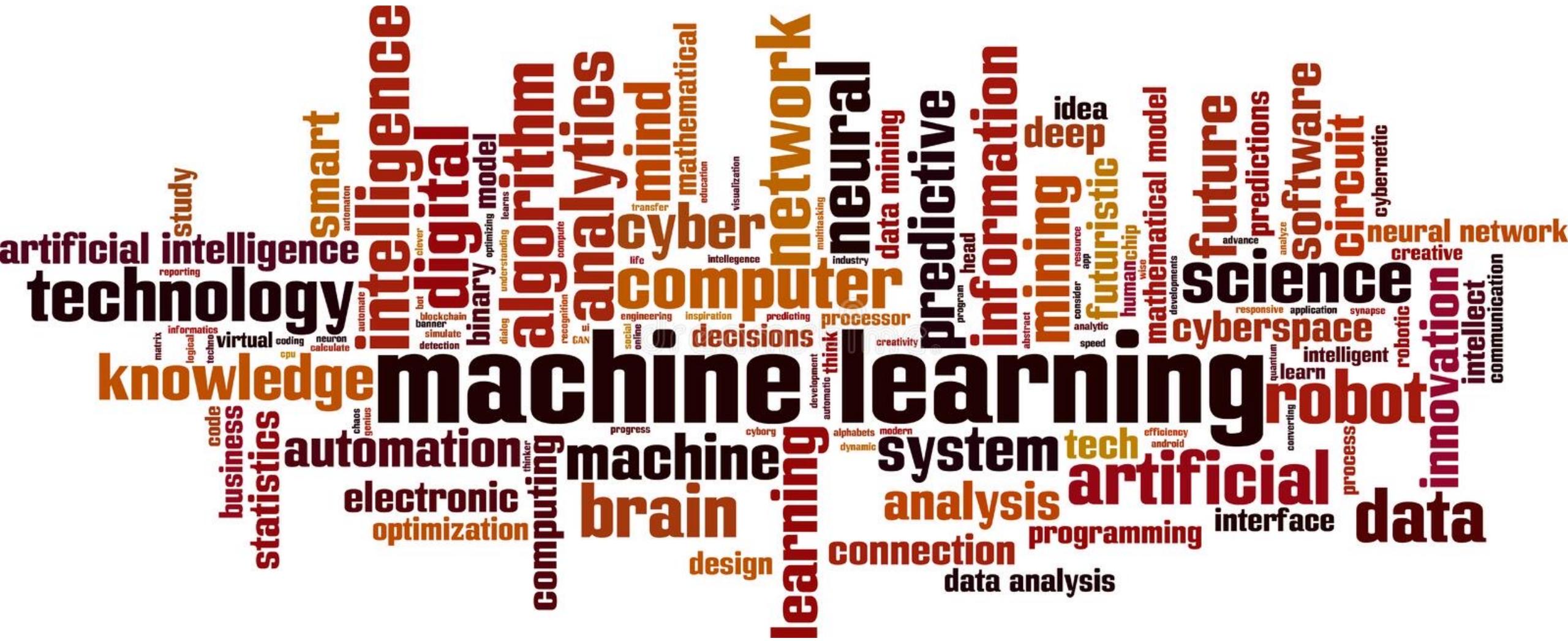
Материалы курса

- ▶ Вики страничка:
[http://wiki.cs.hse.ru/Машинное_обучение_\(сов._бак._ВШЭ-РЭШ_2022\)](http://wiki.cs.hse.ru/Машинное_обучение_(сов._бак._ВШЭ-РЭШ_2022))
 - материалы лекций и семинаров
 - ссылки на доп. курсы и книги
- ▶ Чат в ТГ
- ▶ Лекции по вторникам в 16:20 в R205
- ▶ Семинары
 - по вторникам в 18:10 в D511
 - по средам в 16:20 в

Оценки

- ▶ 4 – 5 домашних заданий
 - Практические задачи на программирование
 - Мягкий и жесткий дедлайны
- ▶ Контрольная работа в ноябре
 - Теоретические вопросы + задачи
- ▶ Экзамен в конце курса
 - Теоретические вопросы + задачи
- ▶ Правила выставления оценок:
$$\text{Итоговая} = \text{Округление} (0.4 * \text{ДЗ} + 0.3 * \text{КР} + 0.3 * \text{Э})$$

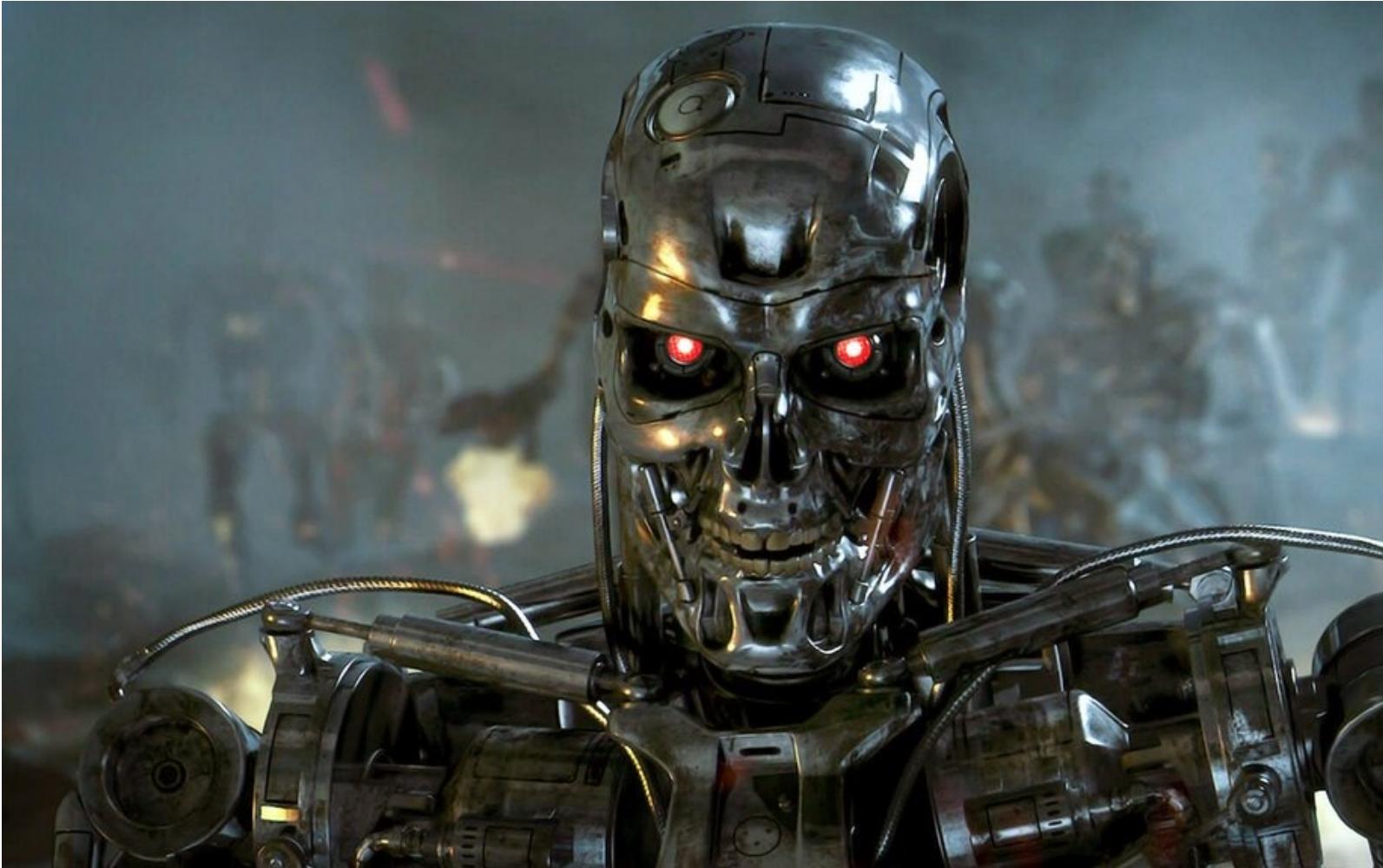




ИИ в кино



Терминатор



Я, робот



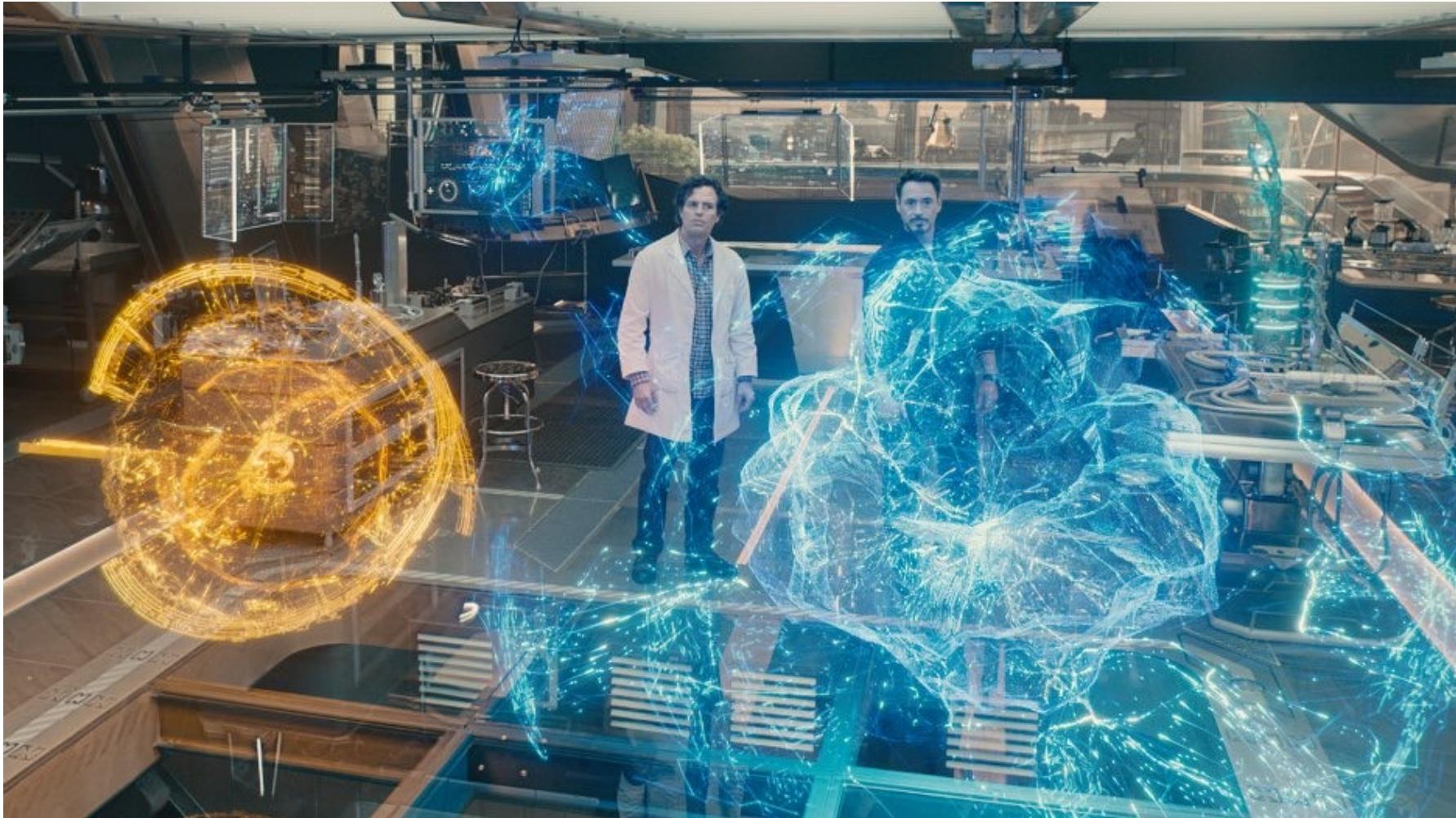
ВАЛЛ-И



Превосходство



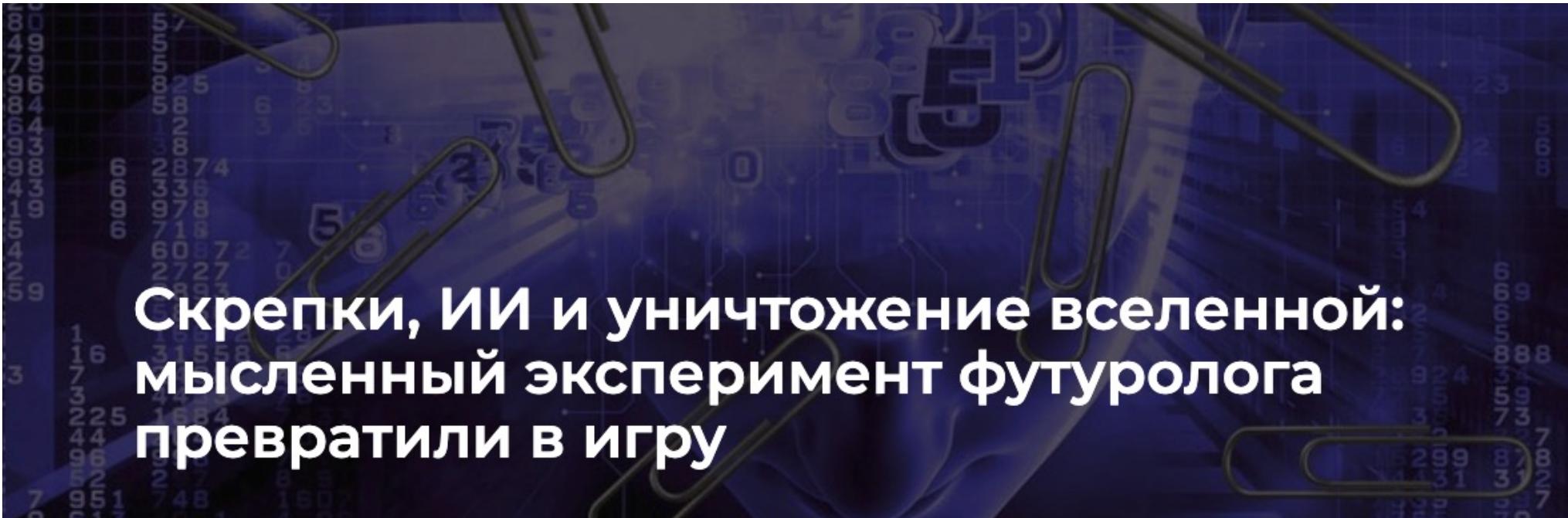
Железный человек / Мстители



Индекс страха



ИИ и скрепки

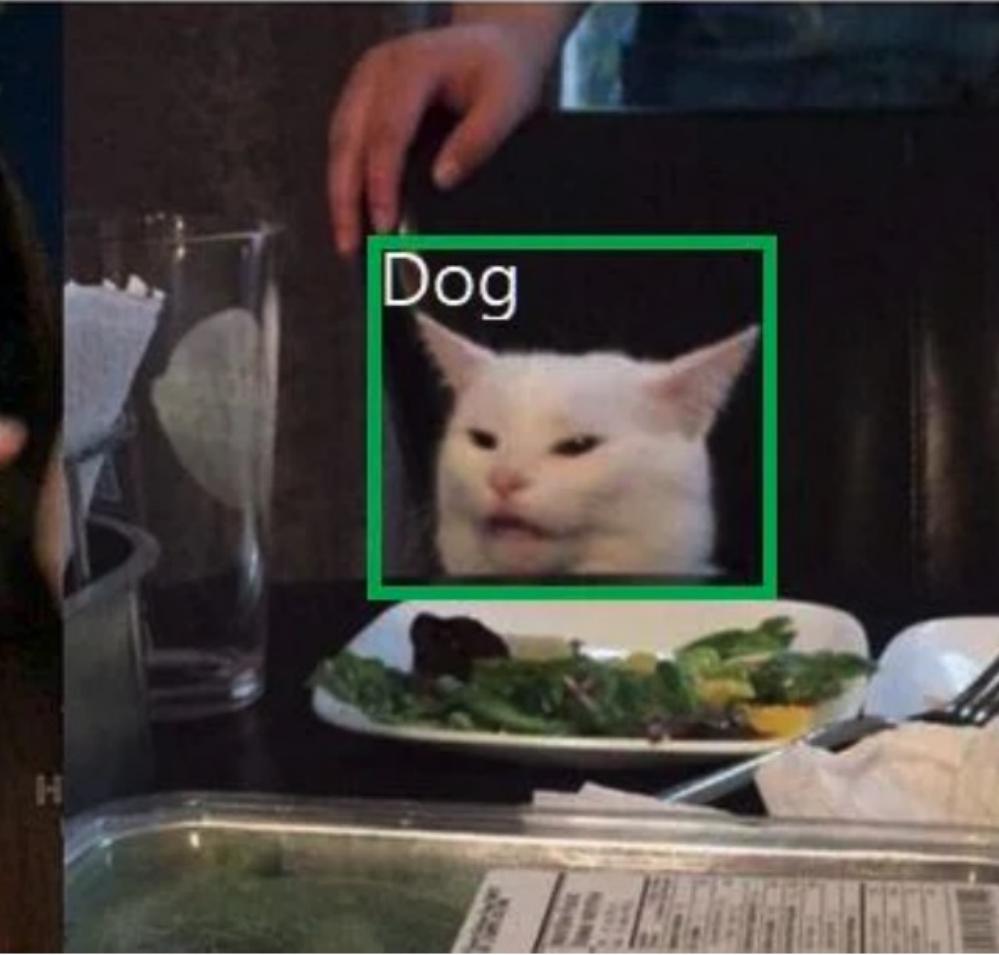


- ▶ Статья: <https://rb.ru/story/universal-paperclips/>
- ▶ Игра-кликер: <http://www.decisionproblem.com/paperclips/index2.html>

People with no idea
about AI, telling me my
AI will destroy the world



Me wondering why my
neural network is
classifying a cat as a dog..



Реальные приложения



Поиск информации

Google machine learning

All Images Videos News Books More Tools

About 2,080,000,000 results (0.54 seconds)

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

<https://www.techtarget.com/searchenterpriseai/definition> :: [What Is Machine Learning and Why Is It Important? - TechTarget](#)



The diagram shows 'Machine Learning' at the center, connected to three main branches: 'Unsupervised Learning' (blue), 'Supervised Learning' (red), and 'Reinforcement Learning' (orange). Each branch has associated examples: Unsupervised includes 'Image Segmentation', 'Customer Segmentation', and 'Anomaly Detection'; Supervised includes 'Image Classification', 'Natural Language Processing', and 'Predictive Maintenance'; Reinforcement includes 'Autonomous Driving', 'Robotics', and 'Gaming'.

About featured snippets · Feedback

https://en.wikipedia.org/wiki/Machine_learning :: [Machine learning - Wikipedia](#)

Machine learning (ML) is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve ...

[Machine Learning \(journal\)](#) · [Machine learning control](#) · [Active learning \(machine...](#)

See results about

machine learning Dictionary definition >

People also ask :

What is machine learning with example? ▾

What are the 3 types of learning in machine learning? ▾

What is AI vs machine learning? ▾

Why is machine learning used? ▾

Feedback

Голосовые помощники

The screenshot shows the Yandex.Alice mobile application. At the top center is the Alice logo, a stylized white egg inside a purple circle. Below it is a large white speech bubble containing the text "Привет, я Алиса!". The background is a solid purple color. In the center, the text "Я готова помочь" is displayed in white. Below this, there are six horizontal cards, each consisting of an icon and text. The cards are arranged in two rows of three. The first row contains: "Определить песню" (Song recognition), "Узнать, что на фото" (Identify what's in the photo), and "Включить сказку" (Play a story). The second row contains: "Одеться по погоде" (Dress according to the weather), "Поиграть" (Play), and "Построить маршрут" (Build a route). The third row contains: "Вызвать такси" (Call a taxi), "Найти нужное место" (Find the right place), and "Купить на Беру" (Buy on Beru). Each card has a small arrow pointing to its right.

- Определить песню > Узнать, что на фото > Включить сказку >
- Одеться по погоде > Поиграть > Построить маршрут >
- Вызвать такси > Найти нужное место > Купить на Беру >

Машинный перевод

Яндекс Браузер

Яндекс.Браузер обновился. Версия 21.8.2

1

Hi! I'm David and I lead the NLP team at Yandex

Закадровый перевод видео с английского

Нейросети Яндекса научились сами переводить и озвучивать видео на английском языке. Пока — не везде, но уже скоро любой ролик на английском можно будет смотреть на русском.

Сразу попробовать новую функцию можно [по ссылке](#).

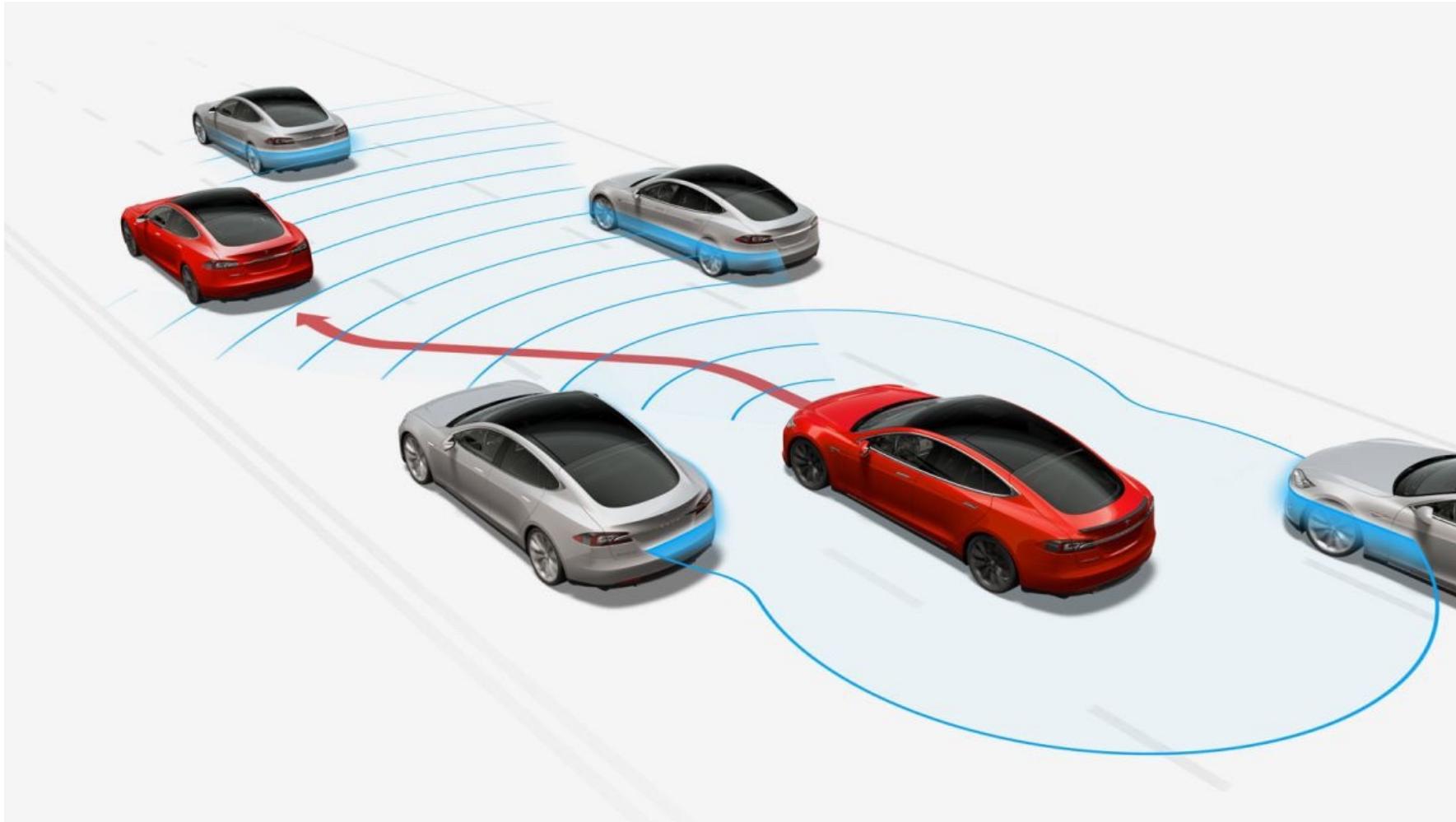
Как включить перевод видео?

Подписывайтесь на новости Яндекс.Браузера:

Дзен ВКонтакте Твиттер Телеграм

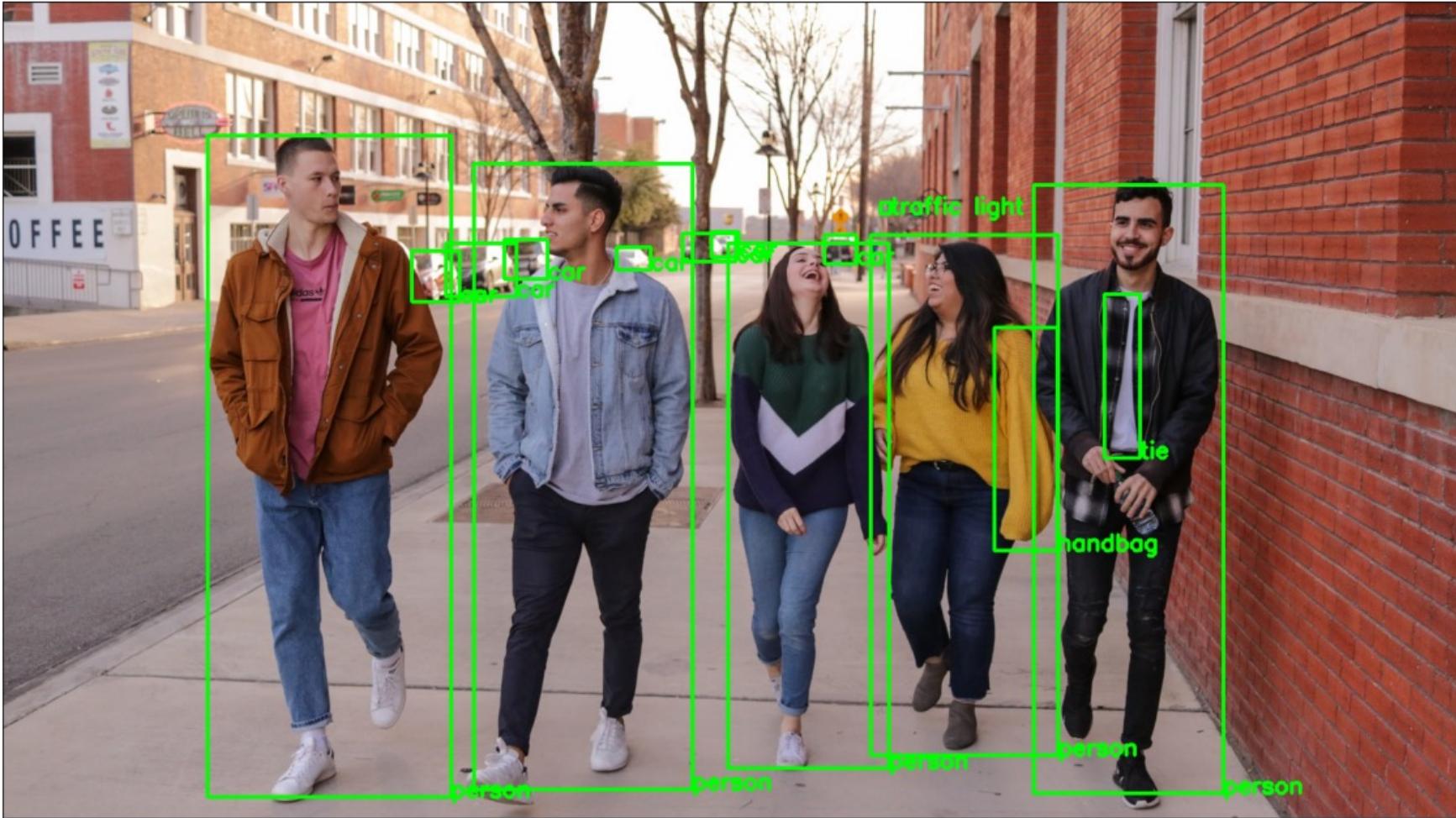
Перевод не доступен для видео, у которых есть технические средства защиты авторских прав (DRM).

АВТОПИЛОТЫ



<https://moscowteslaclub.ru>

Анализ видео



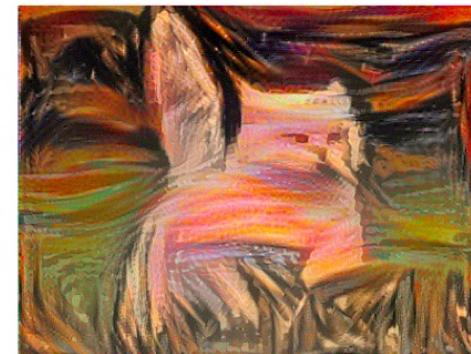
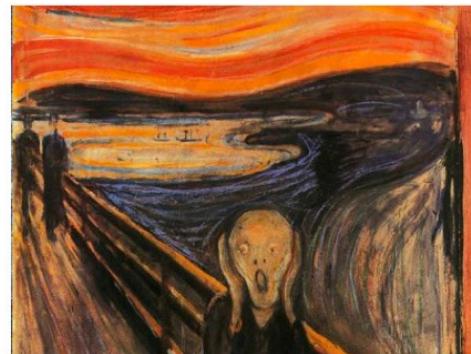
Перенос стиля изображения



+



+



+



Deepfake



<https://www.youtube.com/c/CtrlShiftFace>

АЗАМАТ
МУСАГАЛИЕВ

БРЮС
УИЛЛИС

НАГАРНИКИ

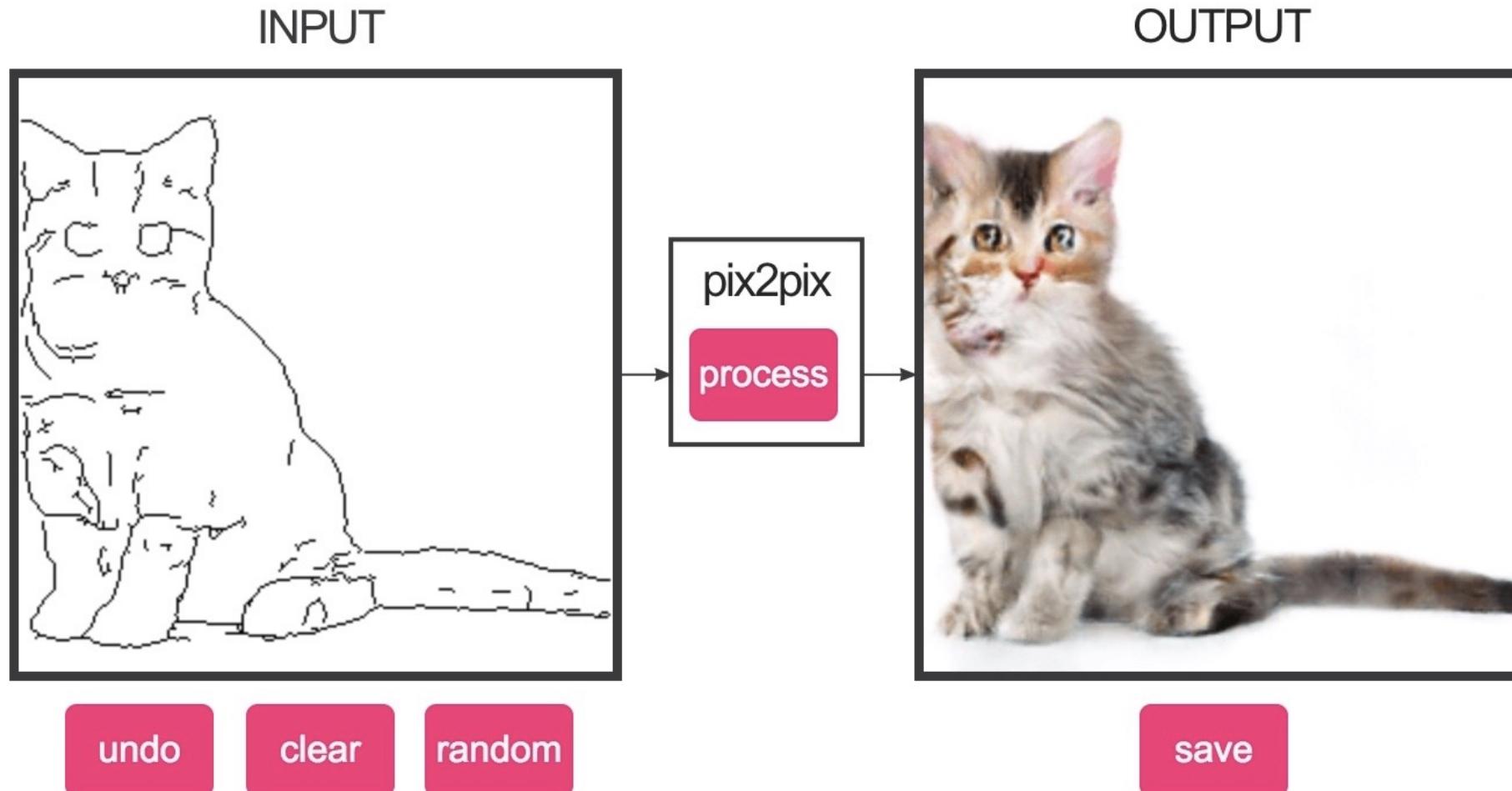
ПЕРВЫЙ СЕМЕЙНЫЙ БЛОКБАСТЕР

С 16 АВГУСТА

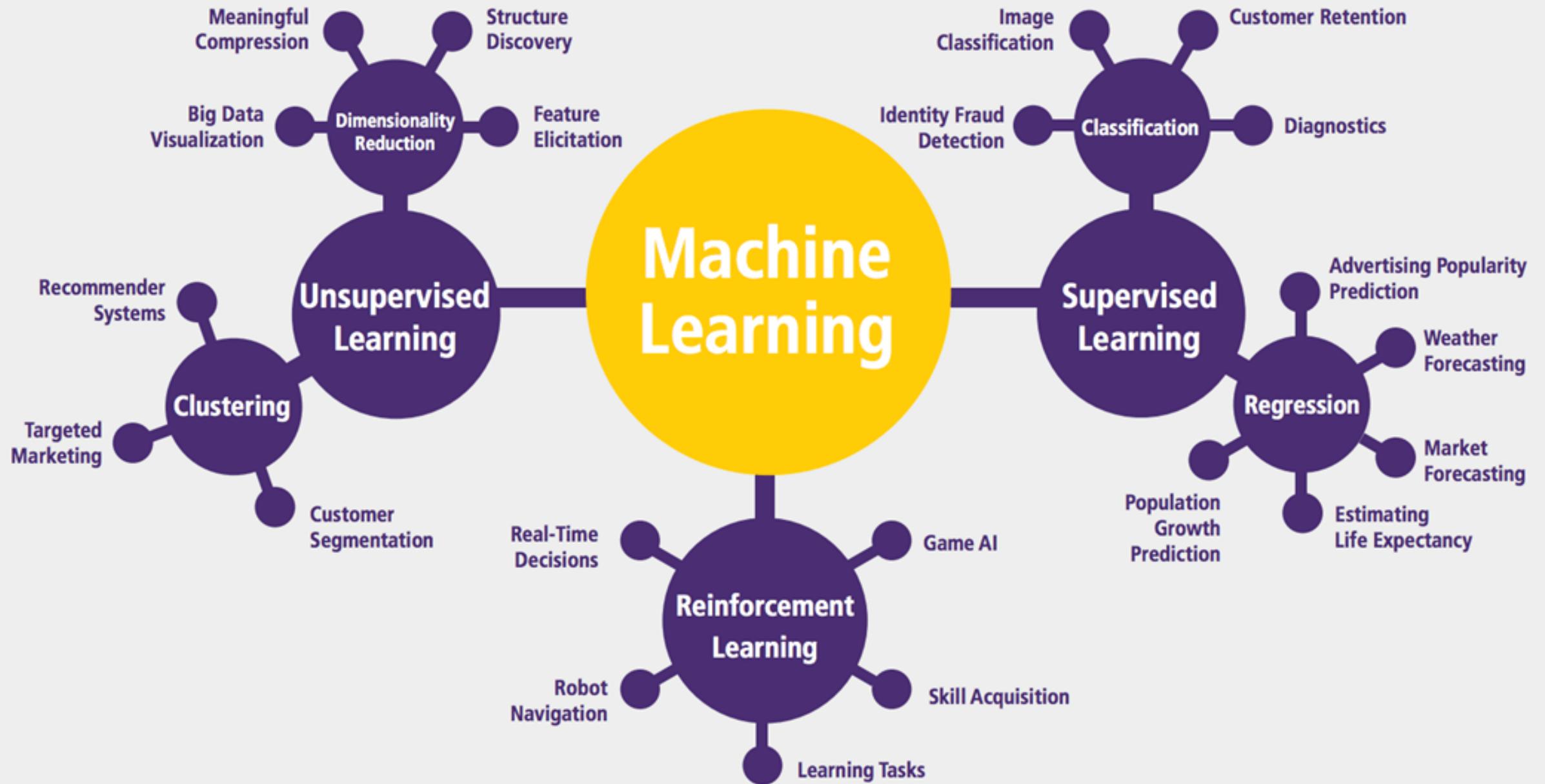


МЕГАФОН
представляет

Image-to-Image Translation

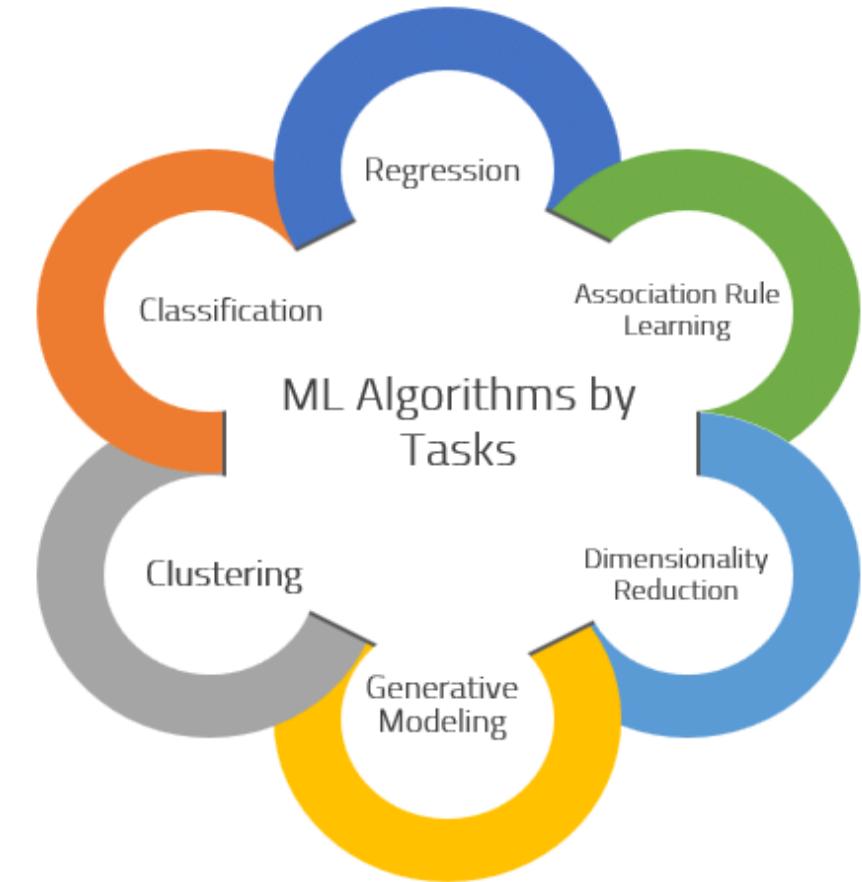


<https://affinelayer.com/pixsrv/>



В нашем курсе

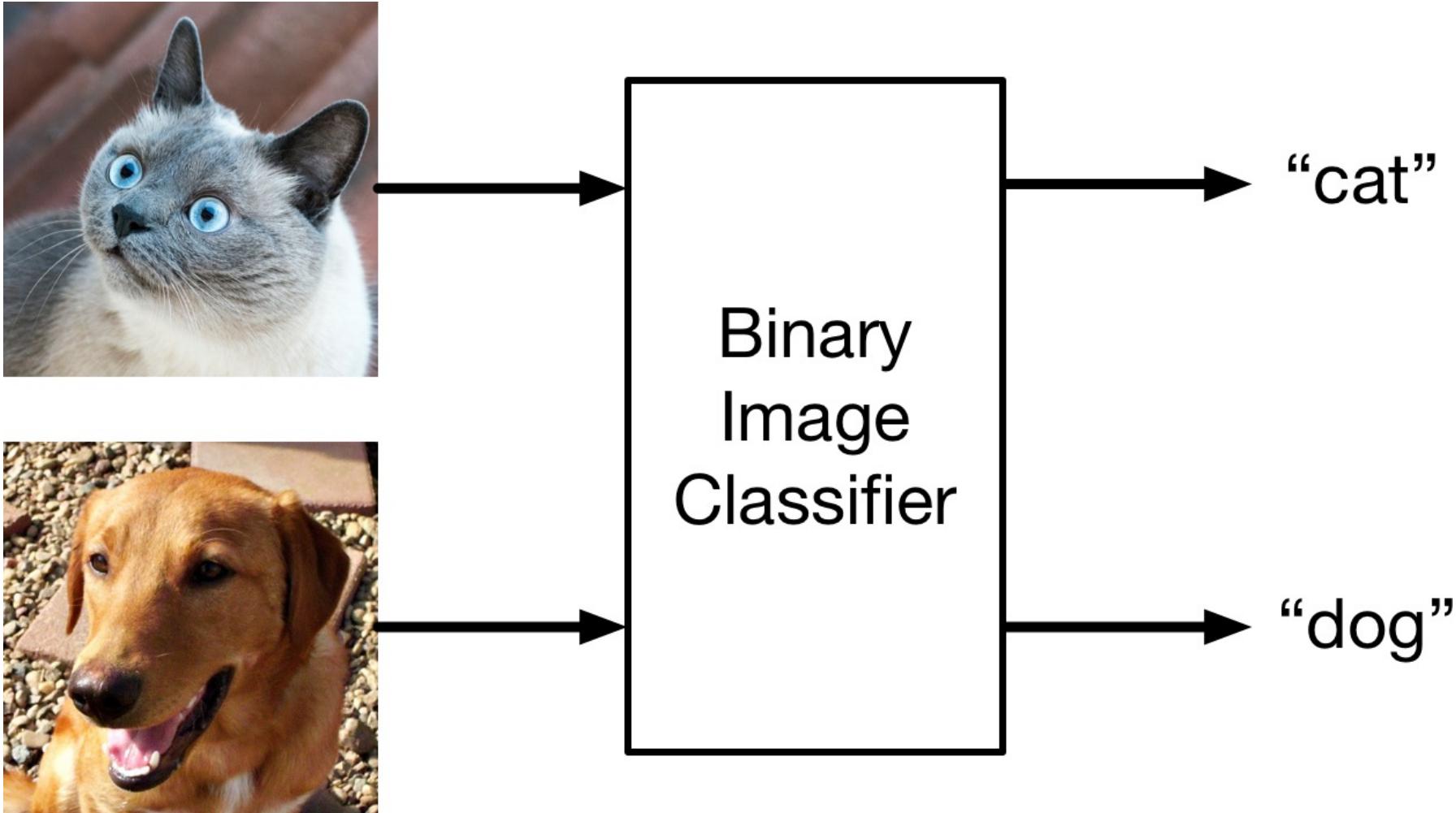
- ▶ Задачи машинного обучения с учителем:
 - Классификация
 - Регрессия
 - Рекомендательные системы
- ▶ Задачи машинного обучения без учителя:
 - Кластеризация
 - Понижение размерности



Классификация



Классификация изображений



<https://www.raywenderlich.com/>

Приложения в банках

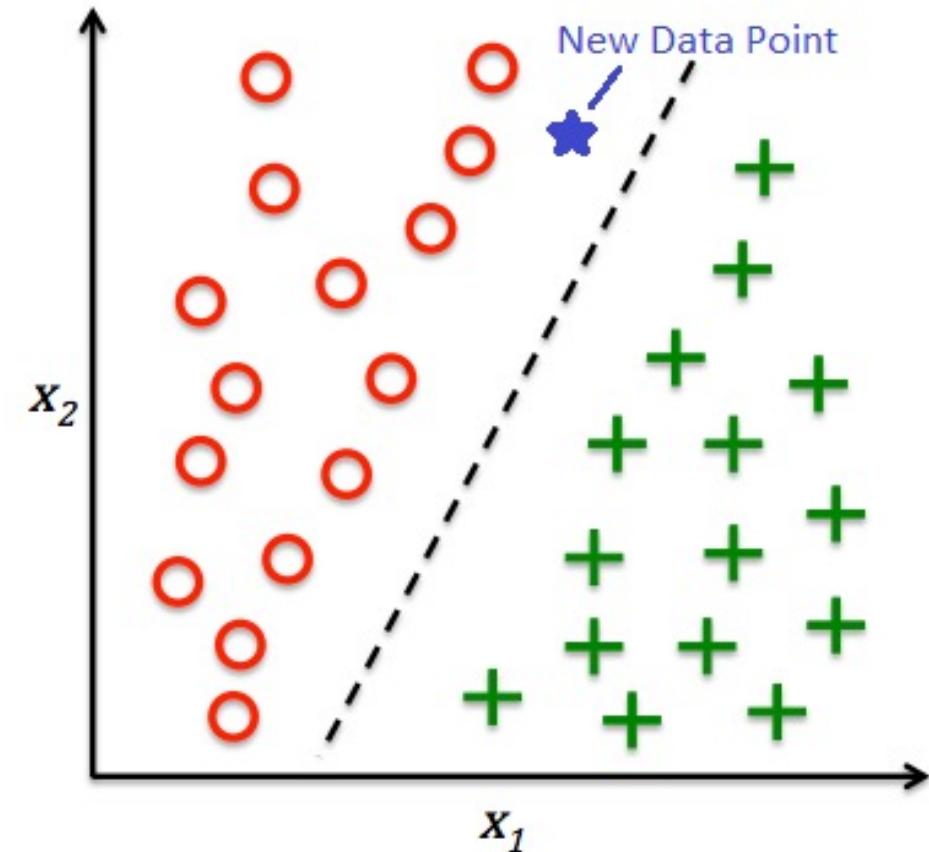
- ▶ Кредитный scoring
- ▶ Прогноз дефолта клиента
- ▶ Обнаружение мошенничества
- ▶ Классификация транзакций



<https://datastart.ru/>

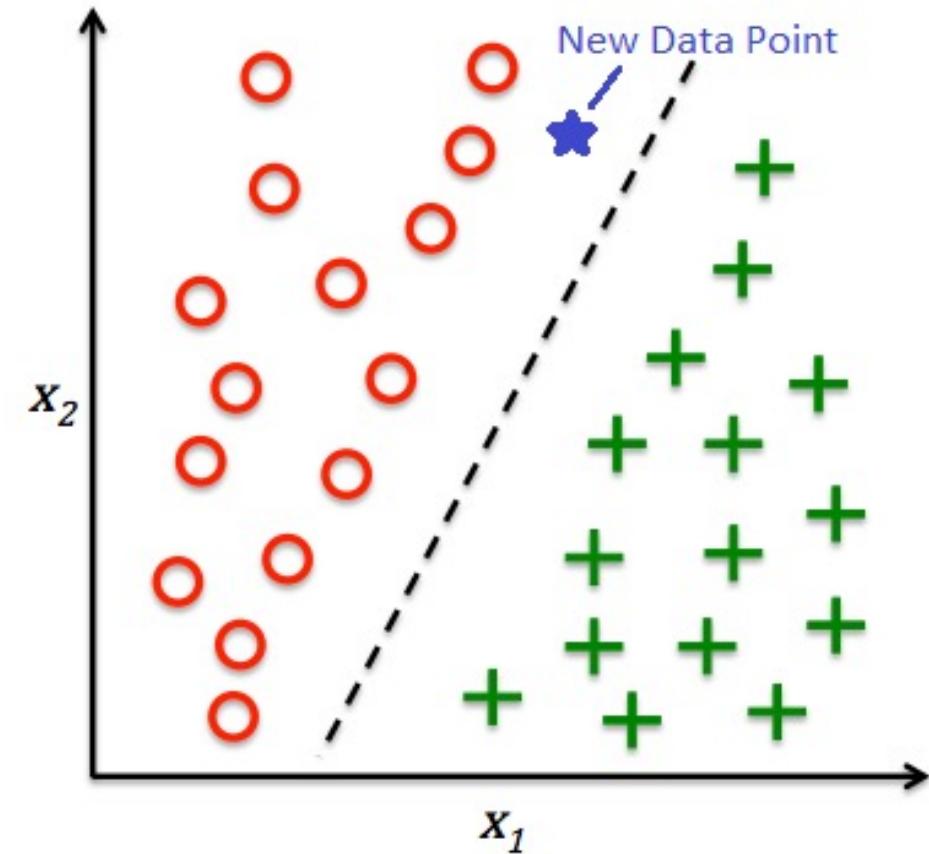
Задача классификации

- ▶ Есть объекты двух (или нескольких) классов
- ▶ Нужно разделить объекты по классам некоторой поверхностью
- ▶ Эту разделяющую поверхность будем называть **классификатором**



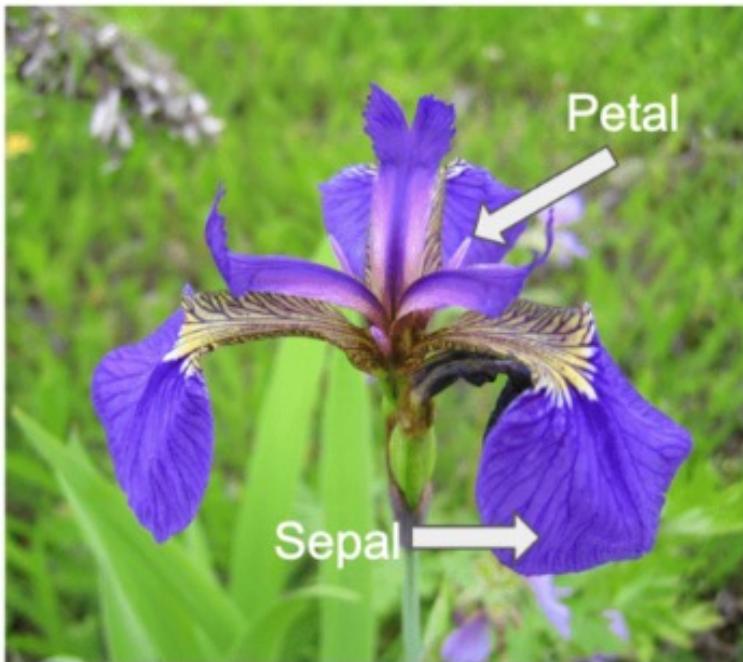
Задача классификации

- ▶ Данные:
 - $X \in \mathbb{R}^{n \times d}$ – матрица объектов;
 - $y \in \{0, 1\}^n$ – вектор меток классов.
 - n – число объектов, d – число признаков
- ▶ Классификатор:
 - $f(\cdot)$ обучаемый классификатор;
 - $y \approx f(X)$

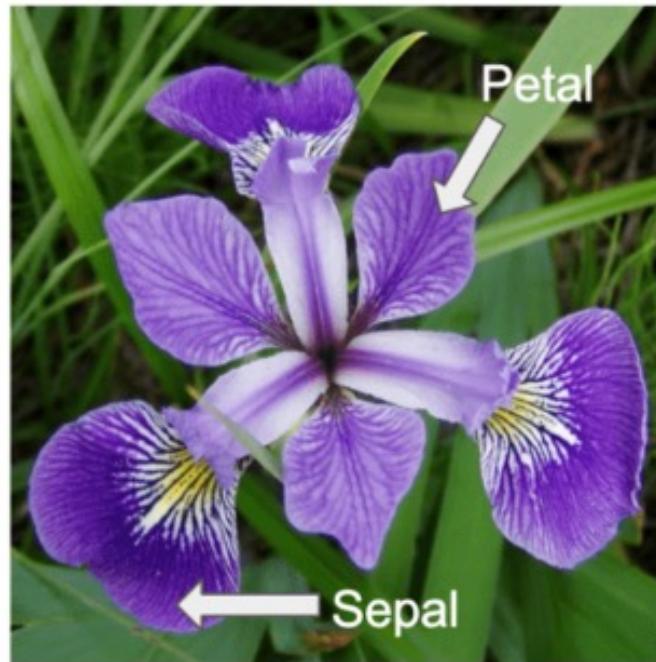


Классификация ирисов

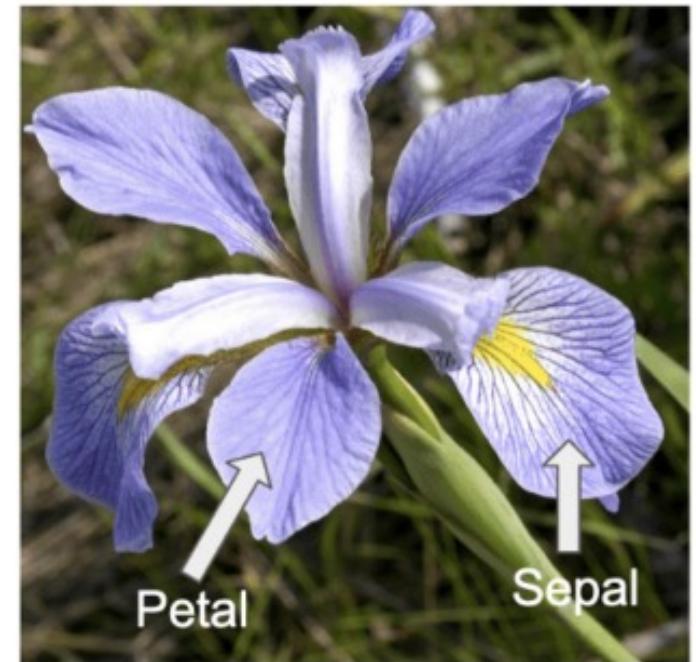
Iris setosa



Iris versicolor



Iris virginica

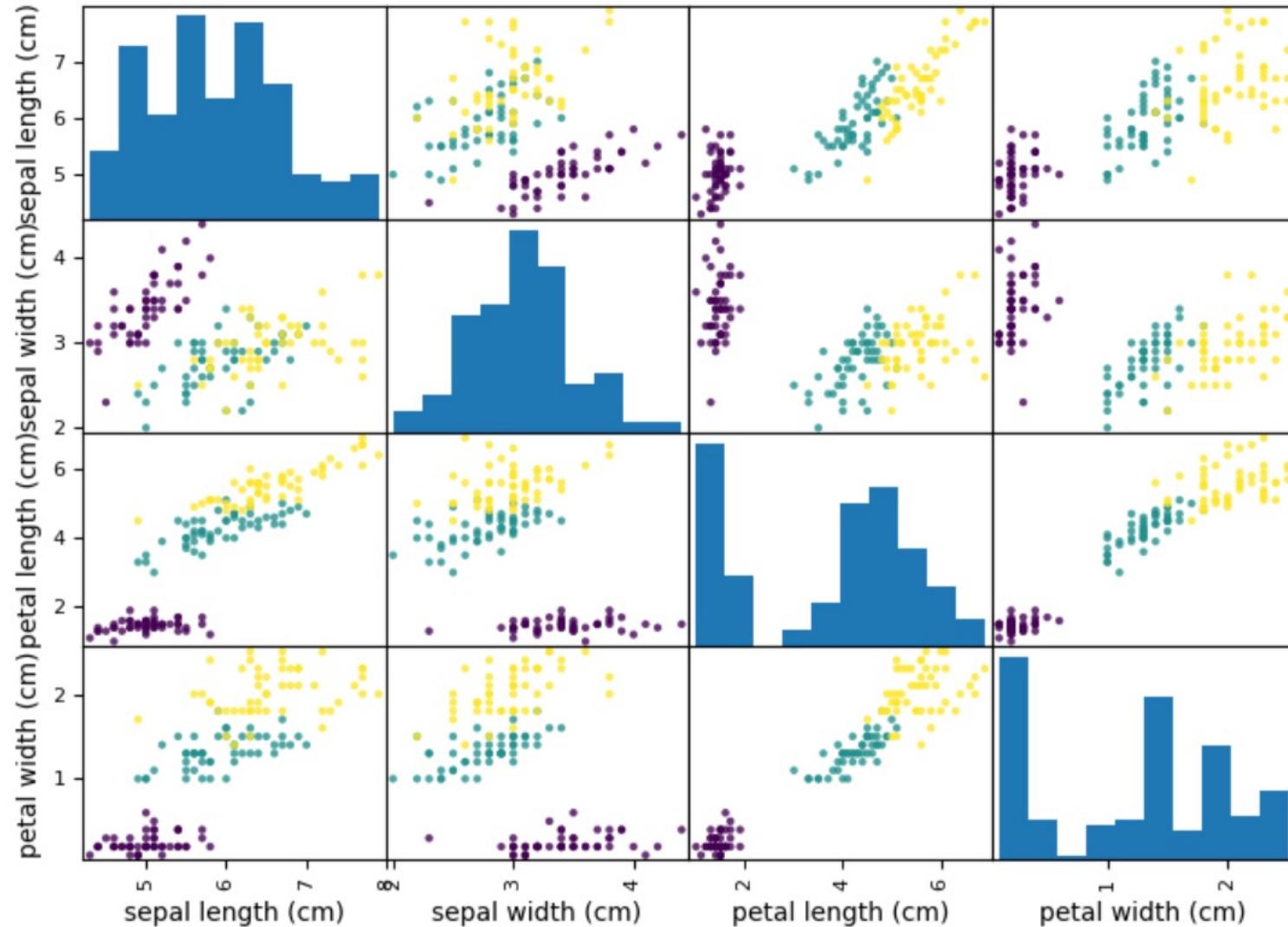


Классификация ирисов

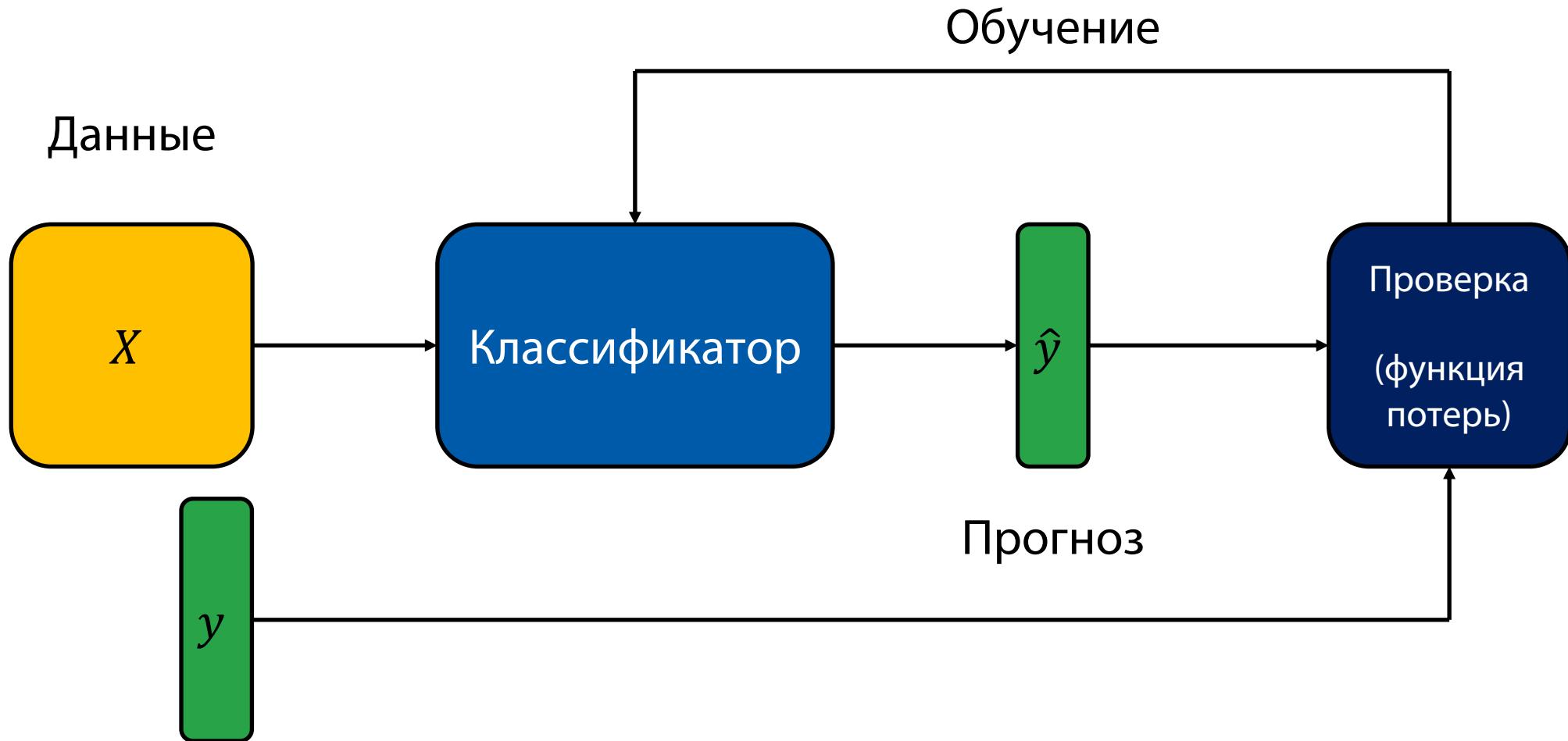
Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	3.6	1.4	0.2	Iris-setosa
6	5.4	3.9	1.7	0.4	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa
8	5	3.4	1.5	0.2	Iris-setosa
9	4.4	2.9	1.4	0.2	Iris-setosa
10	4.9	3.1	1.5	0.1	Iris-setosa
11	5.4	3.7	1.5	0.2	Iris-setosa
12	4.8	3.4	1.6	0.2	Iris-setosa
13	4.8	3	1.4	0.1	Iris-setosa
14	4.3	3	1.1	0.1	Iris-setosa
15	5.8	4	1.2	0.2	Iris-setosa
16	5.7	4.4	1.5	0.4	Iris-setosa
17	5.4	3.9	1.3	0.4	Iris-setosa
18	5.1	3.5	1.4	0.3	Iris-setosa
19	5.7	3.8	1.7	0.3	Iris-setosa



Классификация ирисов



Обучение классификатора



Регрессия



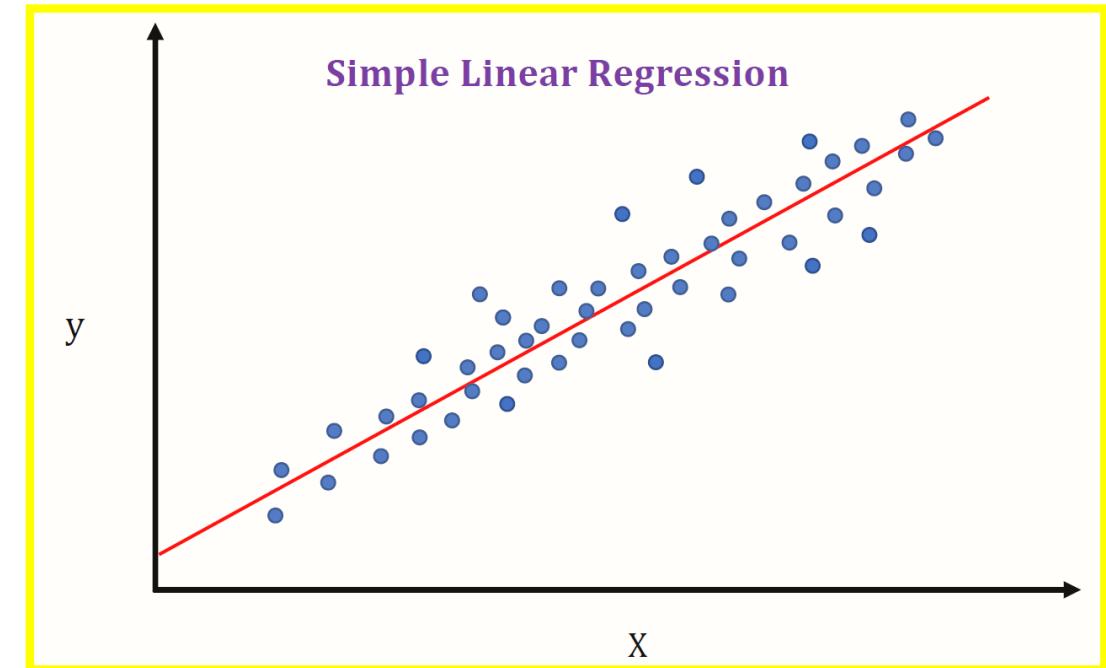
Примеры

- ▶ Прогноз продолжительности жизни
- ▶ Оценка рисков в банках
- ▶ Прогнозирование цены товара
- ▶ Прогнозирование объема продаж



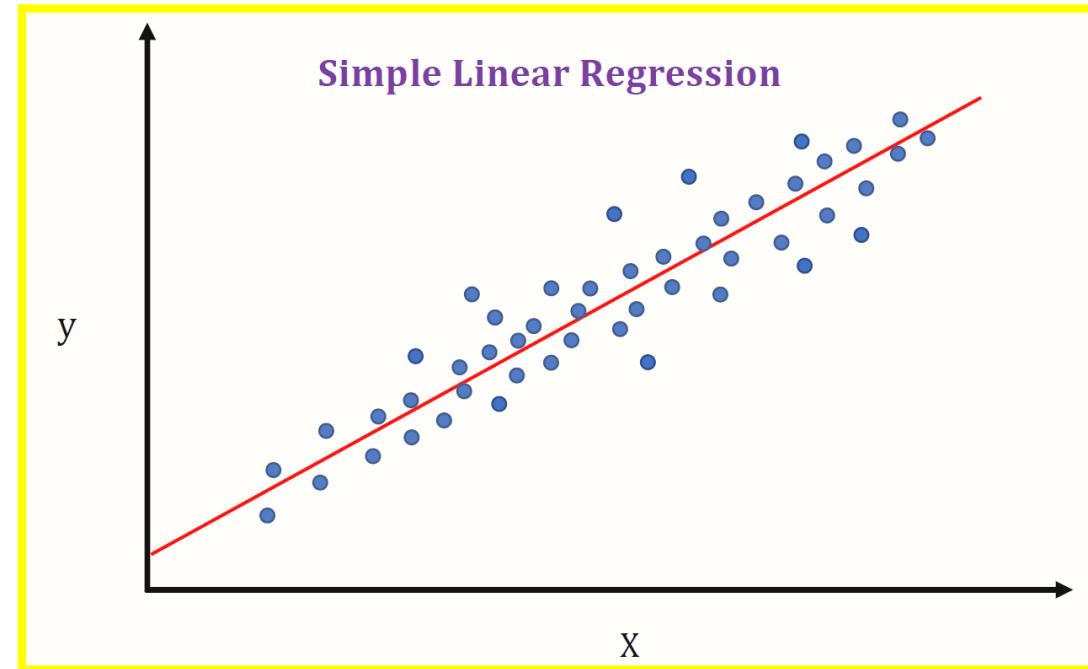
Задача регрессии

- ▶ Есть объекты (X)
- ▶ Нужно предсказать некоторую величину (y)
- ▶ Функция, которая описывает зависимость y от X - **модель регрессии**



Задача регрессии

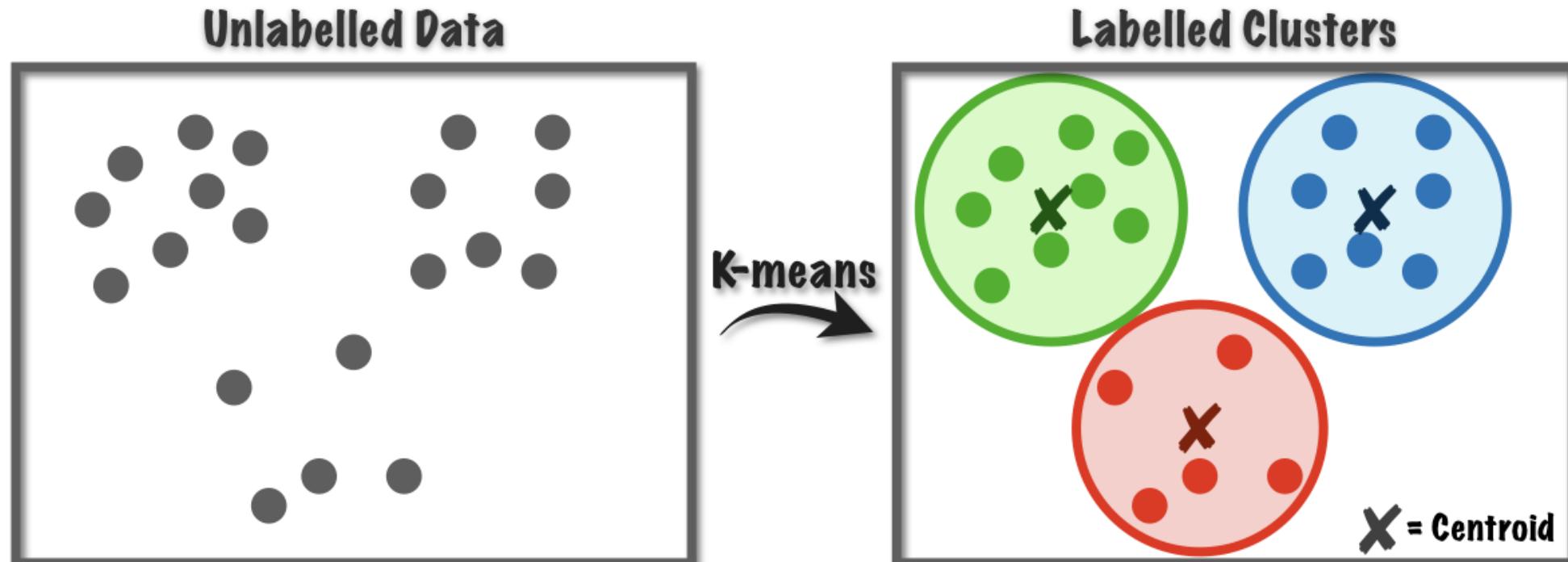
- ▶ Данные:
 - $X \in \mathbb{R}^{n \times d}$ – матрица объектов;
 - $y \in \mathbb{R}^n$ – вектор значений для прогноза.
 - n – число объектов, d – число признаков
- ▶ Регрессор:
 - $f()$ – модель регрессии;
 - $y \approx f(X)$



Кластеризация



Задача кластеризации

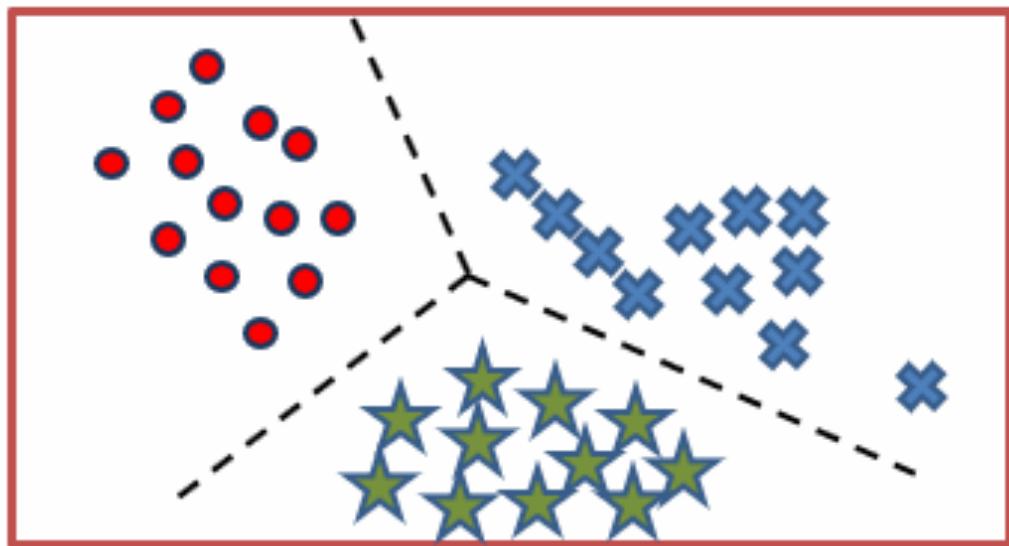


<https://towardsdatascience.com/>

В задаче кластеризации **нет меток классов!**

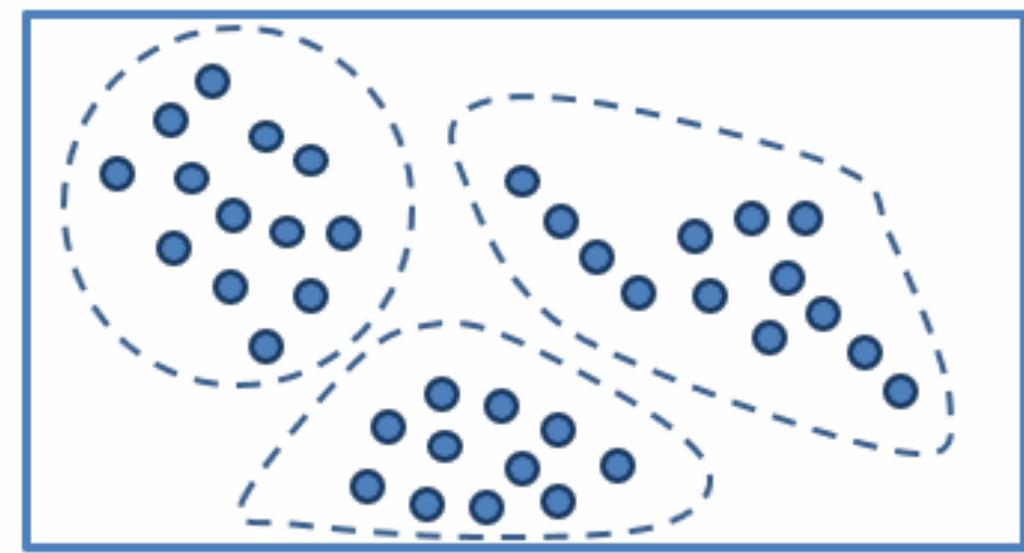
Задача кластеризации

Classification



Supervised learning

Clustering



Unsupervised learning

В задаче кластеризации **нет меток классов!**

Приложения

- ▶ Сегментация клиентов в маркетинге
- ▶ Определение групп пользователей в социальных сетях
- ▶ Кластеризация результатов поиска информации



Рекомендательные системы



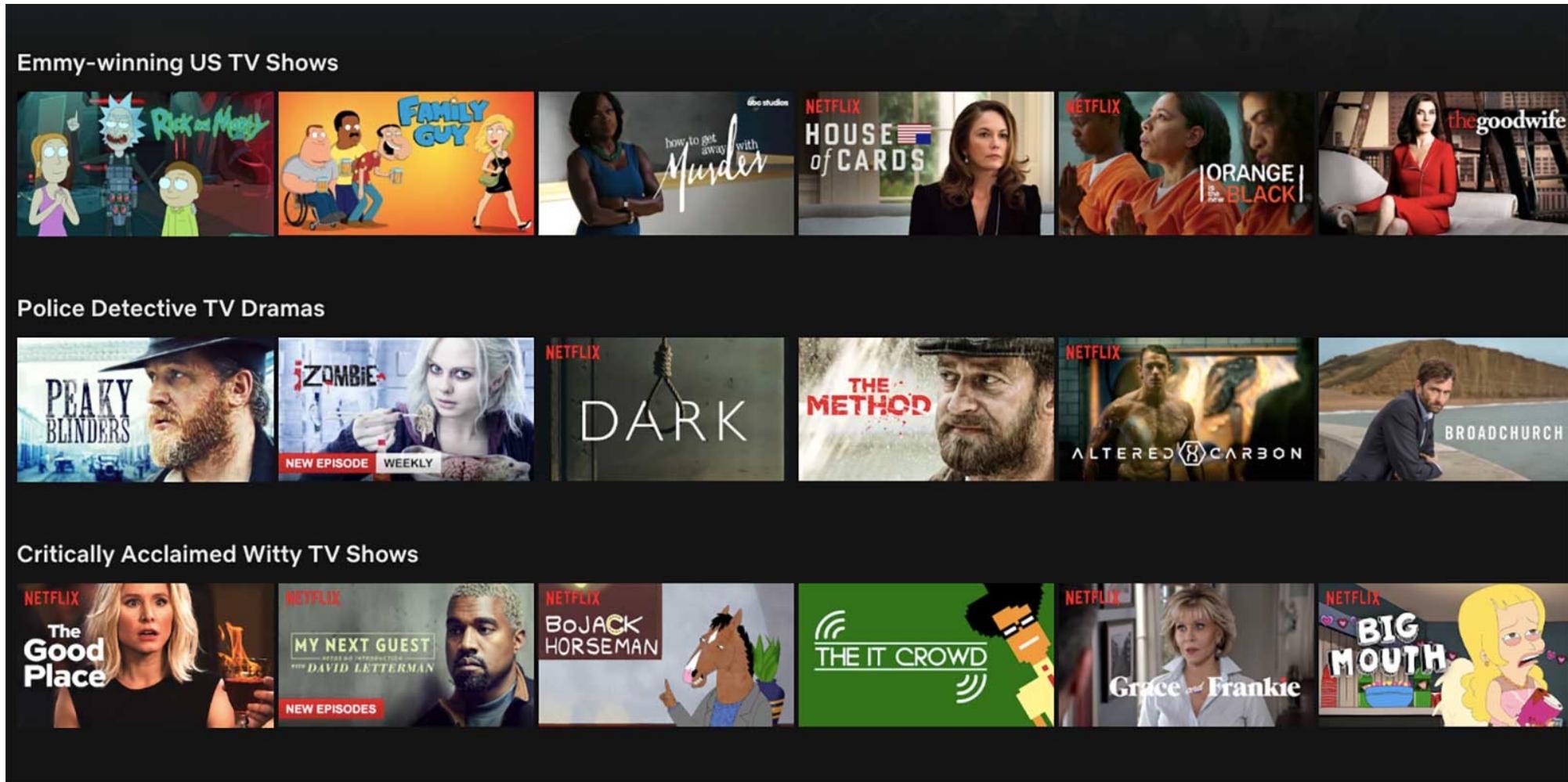
Рекомендательные системы

- ▶ Есть пользователи (users)
- ▶ Есть объекты (items)
- ▶ Есть оценки объектов от пользователей
- ▶ Хотим предсказать оценку пользователя для объекта

	Book 1	Book 2	Book 3	Book 4	Book 5
User A	+	-	+	+	+
User B		+		-	-
User C	+	+	-		
User D		+	?		-

<https://www.advancinganalytics.co.uk/>

Рекомендация фильмов



Рекомендация товаров

OZON
РАБОТАЕМ ВСЕГДА

Каталог

Везде ▾ Искать на Ozon

Хочу Premium Аутлет! Акции Супермаркет Бренды Магазины Книги Электроника Одежда и обувь Детские товары Дом и сад Ozon Card Авиа и ж/д билеты

Светлана Заказы Избранное Корзина 44

Вернем 5% кэшбэка на все товары Ozon



Избранное

Списки товаров Сначала новые Редактировать

Все товары Мой список Книги Новый список Все списки

Только закончился Только закончился Скидка 500 р.



2 490 ₽

Беспроводные наушники JBL Type 500, черный

★★★★★ 134 отзыва

Похожие



283 ₽ 365 ₽

Цена с Premium ниже

Специальное чистящее средство Aware Mommy 903-HP014

★★★★★ 1 отзыва

Похожие



3 990 ₽ 4 988 ₽

Бестселлер

Пароварка Tefal VC 1301 Minicompact

★★★★★ 99 отзывов

В корзину



239 ₽

Бестселлер

Средство для чистки деревянных поверхностей "Emsal", 750 мл

★★★★★ 22 отзыва

В корзину

Наличие товаров

Неважно В наличии Не в наличии

Все категории товаров

Одежда, обувь и аксессуары Женщинам

Электроника Наушники и аудиотехника

Дом и сад

OZON, доставка со склада OZON

OZOON, доставка со склада OZON

OZON, доставка со склада OZON

Заключение



Резюме лекции

- ▶ Задачи машинного обучения с учителем: данные $X \in \mathbb{R}^{n \times d}$, $y \in \mathbb{R}^n$
 - Регрессия: $y \in \mathbb{R}^n$
 - Классификация:
 - Бинарная: $y \in \{0, 1\}^n$
 - Многоклассовая: $y \in \{0, 1, 2, \dots, c\}^n$
 - Рекомендательные системы
- ▶ Задачи обучения без учителя: данные $X \in \mathbb{R}^{n \times d}$
 - Кластеризация: $X \in \mathbb{R}^{n \times d} \rightarrow Z \in \{0, 1, 2, \dots, c\}^n$
 - Понижение размерности: $X \in \mathbb{R}^{n \times d} \rightarrow Z \in \mathbb{R}^{n \times k}, k < d$