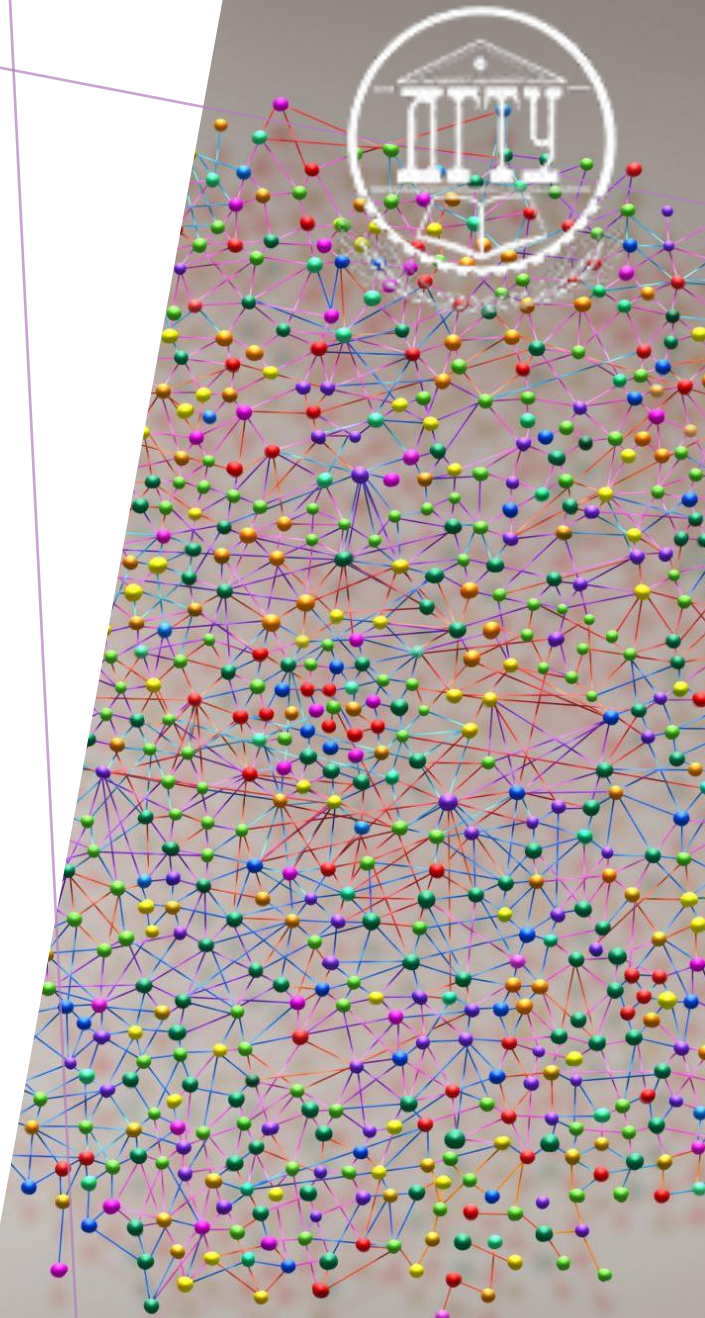
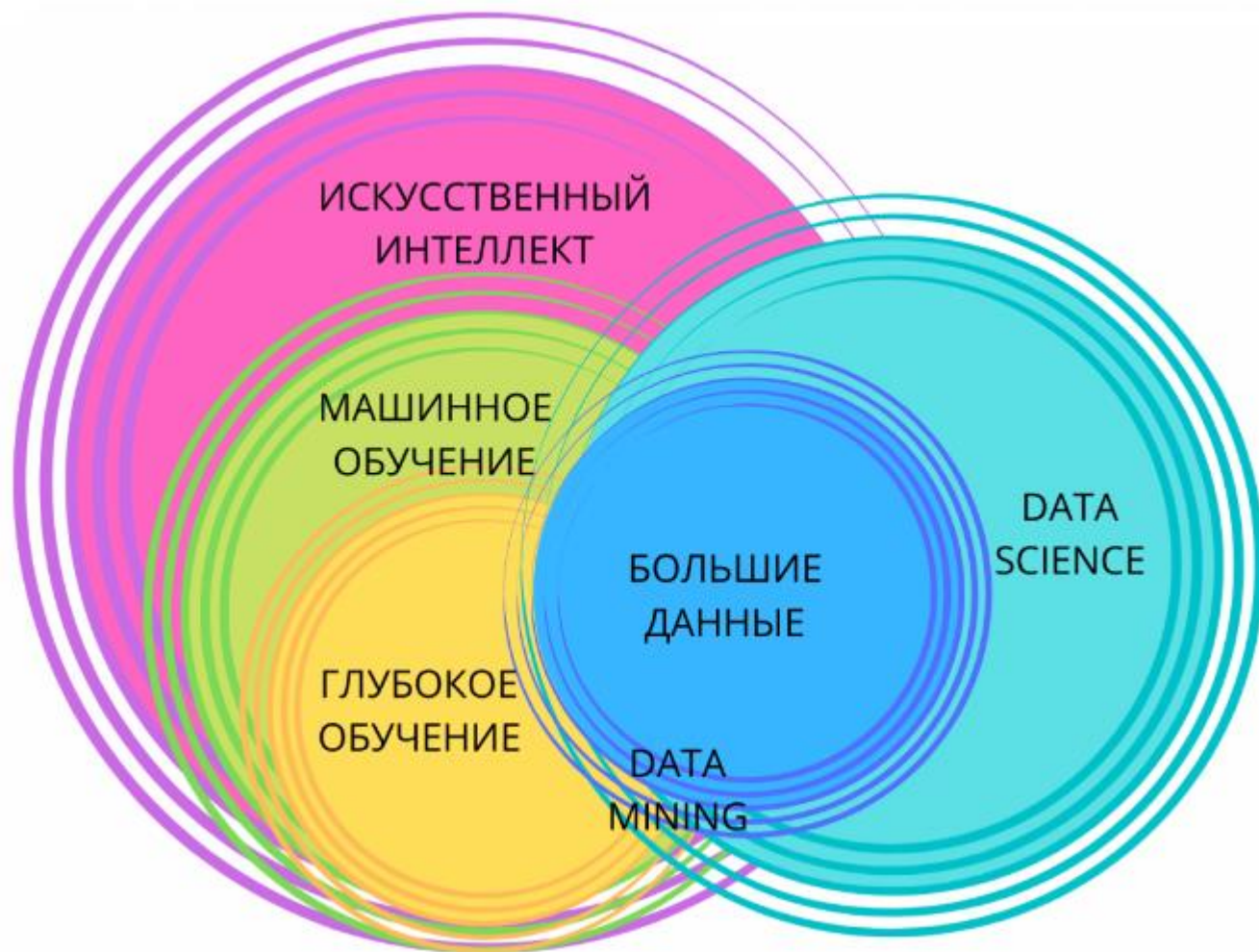


ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

СКЛЯРЕНКО АННА АНАТОЛЬЕВНА



ТЕХНОЛОГИИ ИИ



Data Science

- Это концепция объединения статистики, анализа данных, машинного обучения и связанных с ними методов для понимания и анализа реальных явлений.

Data Mining

- Широкое понятие, означающее извлечение знаний из данных.

КЛЮЧЕВЫЕ ПОНЯТИЯ

- **Искусственный интеллект (ИИ)** — это разум, встроенный в машину.
- **Машинное обучение** — это процесс внедрения интеллекта в систему или машину **без явного программирования**.
- **Глубокое обучение** — алгоритмы основаны на биологической структуре и функционировании мозга и призваны наделить машины интеллектом.



МАШИННОЕ ОБУЧЕНИЕ

- Способность компьютера с системой ИИ **принимать решения по результатам обработки данных**, не придерживаясь четких схем и правил.
- **Машина** становится **способна к поиску закономерностей** в сложных для человека задачах с большим числом параметров, к обнаружению точных ответов и верному прогнозированию.
- В компьютер (или отдельную программу) закладывают **алгоритм самостоятельного нахождения решений** путём комплексного использования статистических данных, из которых выводятся **закономерности** и на основе которых делаются **прогнозы**.

КЛЮЧЕВЫЕ ПОНЯТИЯ

Алгоритмы

- специальные программы для выбора источника данных

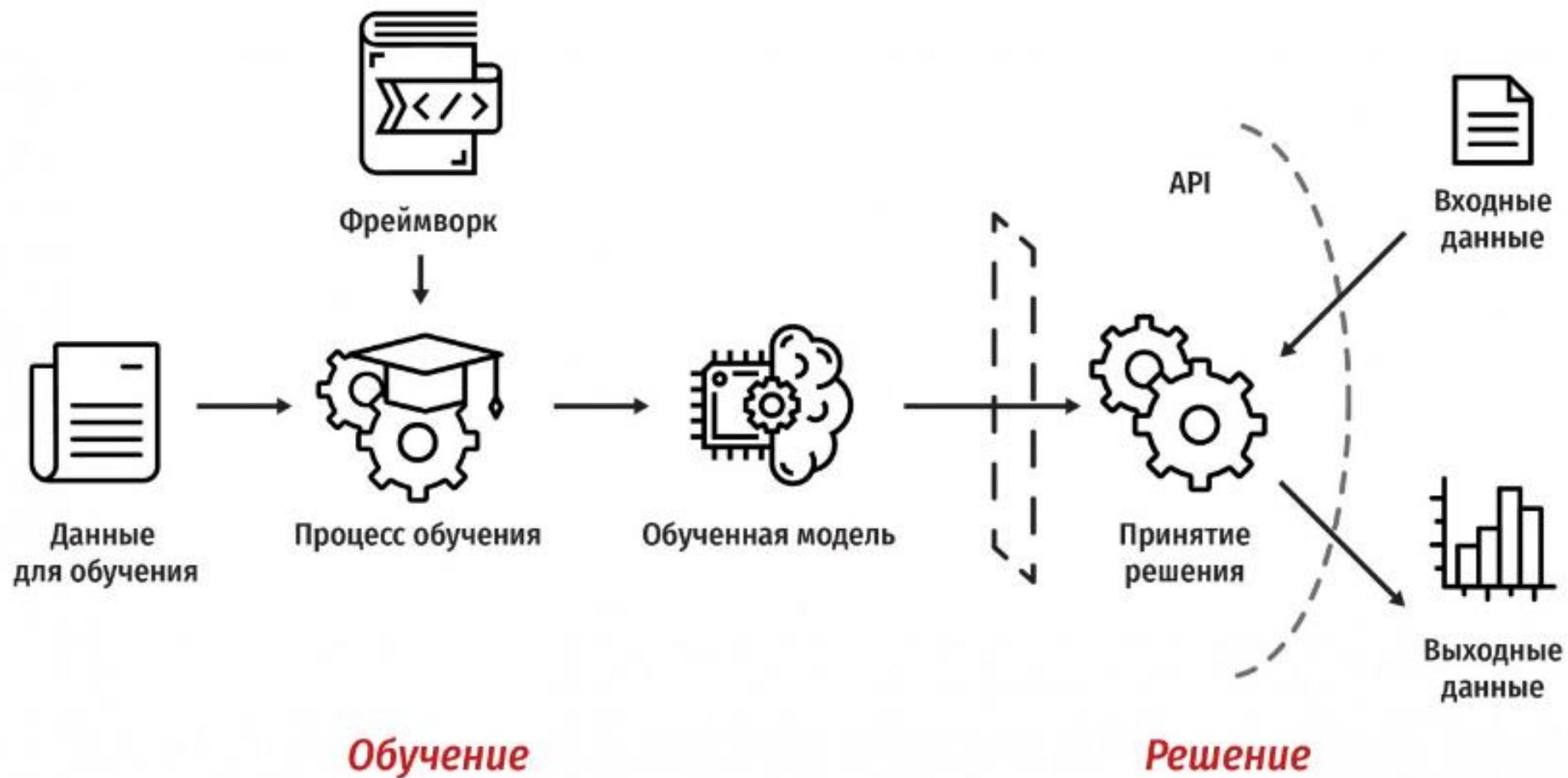
Набор данных

- выборки данных в виде текстовых, графических, видеофайлов, которую машина использует для накопления опыта при обучении.
- для решения каждого конкретного типа задач в систему должны загружаться уникальные данные.

Признаки

- индивидуальные измеримые параметры наблюдаемых явлений, от правильности подбора которых зависит успешность и скорость машинного обучения
- свойства, метрики, фичи, features

ОБЩИЙ ПРОЦЕСС



ТИПЫ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

ЗАДАЧА РЕГРЕССИИ

- прогноз на основе выборки объектов с различными признаками: цена квартиры, стоимость ценной бумаги по прошествии полугода, ожидаемый доход магазина на следующий месяц, качество вина при слепом тестировании.

ЗАДАЧА КЛАССИФИКАЦИИ

- получение категориального ответа на основе набора признаков: есть ли на фотографии кот, является ли изображение человеческим лицом, болен ли пациент раком

ЗАДАЧА КЛАСТЕРИЗАЦИИ

- распределение данных на группы: разделение всех клиентов мобильного оператора по уровню платёжеспособности, отнесение космических объектов к той или иной категории

ЗАДАЧА ВЫЯВЛЕНИЯ АНОМАЛИЙ

- поиск отличий в наборах данных от стандартной информации:
выявление мошеннических действий с банковскими картами.

ЗАДАЧА УМЕНЬШЕНИЯ РАЗМЕРНОСТИ

- сведение большого числа признаков к меньшему: сжатие данных

ОСНОВНЫЕ ВИДЫ МАШИННОГО ОБУЧЕНИЯ



ОБУЧЕНИЕ С УЧИТЕЛЕМ

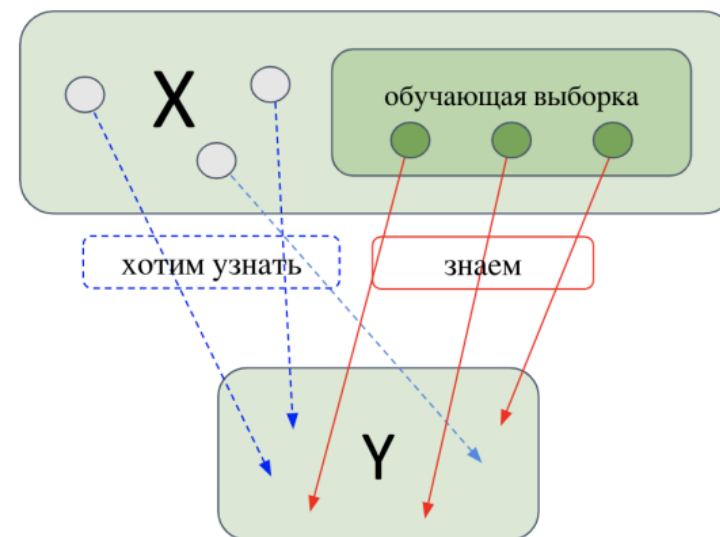
- на вход: набор тренировочных примеров (обучающий или тренировочный набор данных (training set или training sample — тренировочная выборка)
- задача: продолжить уже известные ответы на новый опыт, выраженный обычно в виде тестового набора данных (test set, test sample)

Пример:

- Датасет о квартирах, предсказание рыночной стоимости квартиры по ее параметрам (*задача регрессии*).
- Подтвердить рак у пациента по медицинским показателям (*задача классификации*)

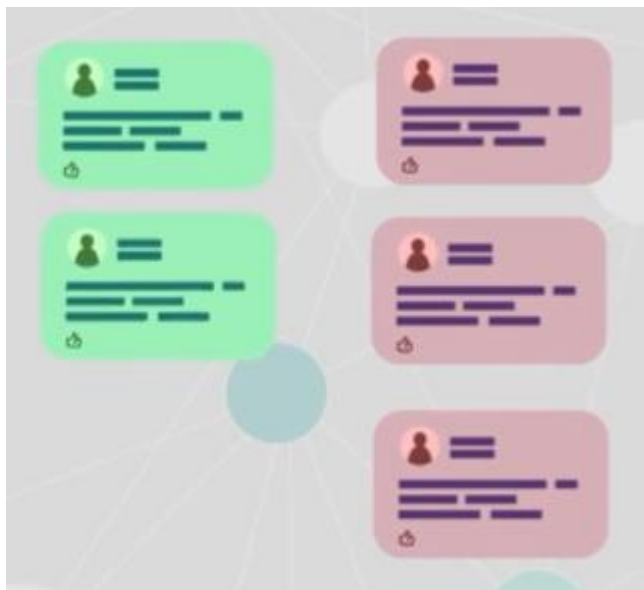
ПОСТАНОВКА ЗАДАЧИ

- ▶ X — множество объектов в пространстве признаков
- ▶ Y — область значений целевой функции
- ▶ $f: X \rightarrow Y$ — неизвестная закономерность (может иметь стохастическую природу!)
- ▶ **Дано:** Обучающая выборка вида $\{(X_i, y_i)\}_{i=1}^n$
- ▶ **Цель:** Максимально точно приблизить f на всём X !



ПРИМЕР: ЛИНЕЙНАЯ КЛАССИФИКАЦИЯ ТЕКСТОВ

Анализ тональности отзывов



Выборка:

Text

Label

Хочу оставить положительный отзыв при посещении банка на улице Менжинского 21 (ДО "Бабушкинский")

1

Очередь не соблюдается! На табло номера которых скорее всего и нет(сидел один....

0

Сегодня в банке была впервые, и осталась очень довольна.
Качество обслуживания на высоком уровне...

1

Позитивный

1

Негативный

0

ПРИМЕР: ЛИНЕЙНАЯ КЛАССИФИКАЦИЯ ТЕКСТОВ

Задача бинарной классификации

Постановка задачи:

- Объект: текст отзыва
- Целевая переменная: {0,1}
- Метрика качества: доля верных ответов

$$\begin{aligned} & [\text{спасибо}] * 10 \\ + & [\text{реклама}] * (-30) \\ + & [\text{повторно}] * (-5) \\ + & [\text{плохо}] * (-10) \\ + & [\text{прекрасно}] * 10 \end{aligned}$$

$$\Sigma = b(x)$$

1. *Подготовка текстов:* разбиение на слова, приведение к начальной форме, удаление частых и редких слов ...
2. *Векторизация (эмбеddинг):* выделение признаков
3. *Построение линейной модели:* $b(x) > 0 \rightarrow 1$, $b(x) < 0 \rightarrow 0$
4. *Проверка модели:* доля ошибок на обучении и на тесте

ОБУЧЕНИЕ БЕЗ УЧИТЕЛЯ

Машина самостоятельно выявляет закономерности, определяет признаки и классифицирует полученные данные. Минимизируют “энтропию” системы.

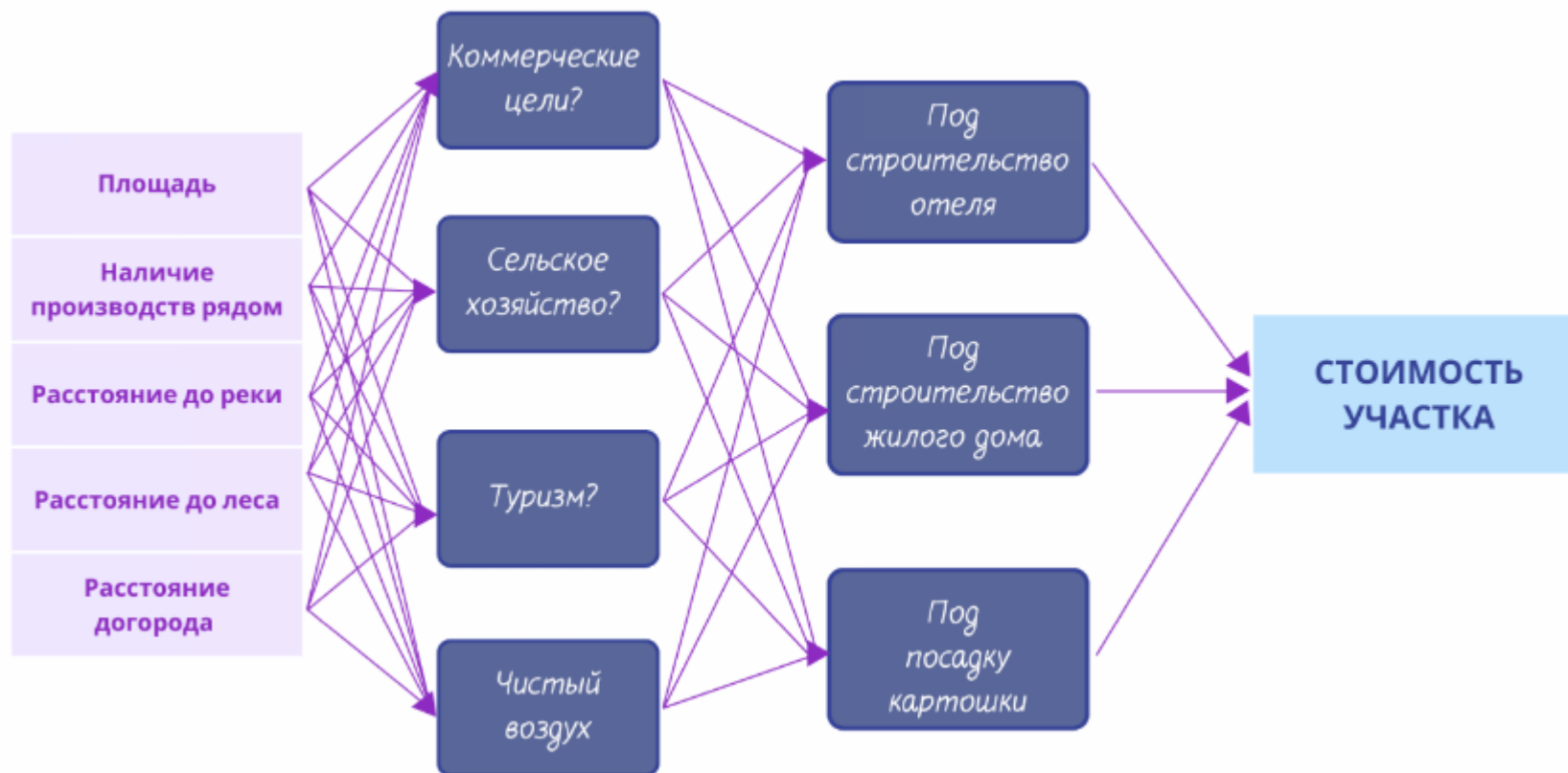
- Задача кластеризации
- Задача снижение размерности

Пример

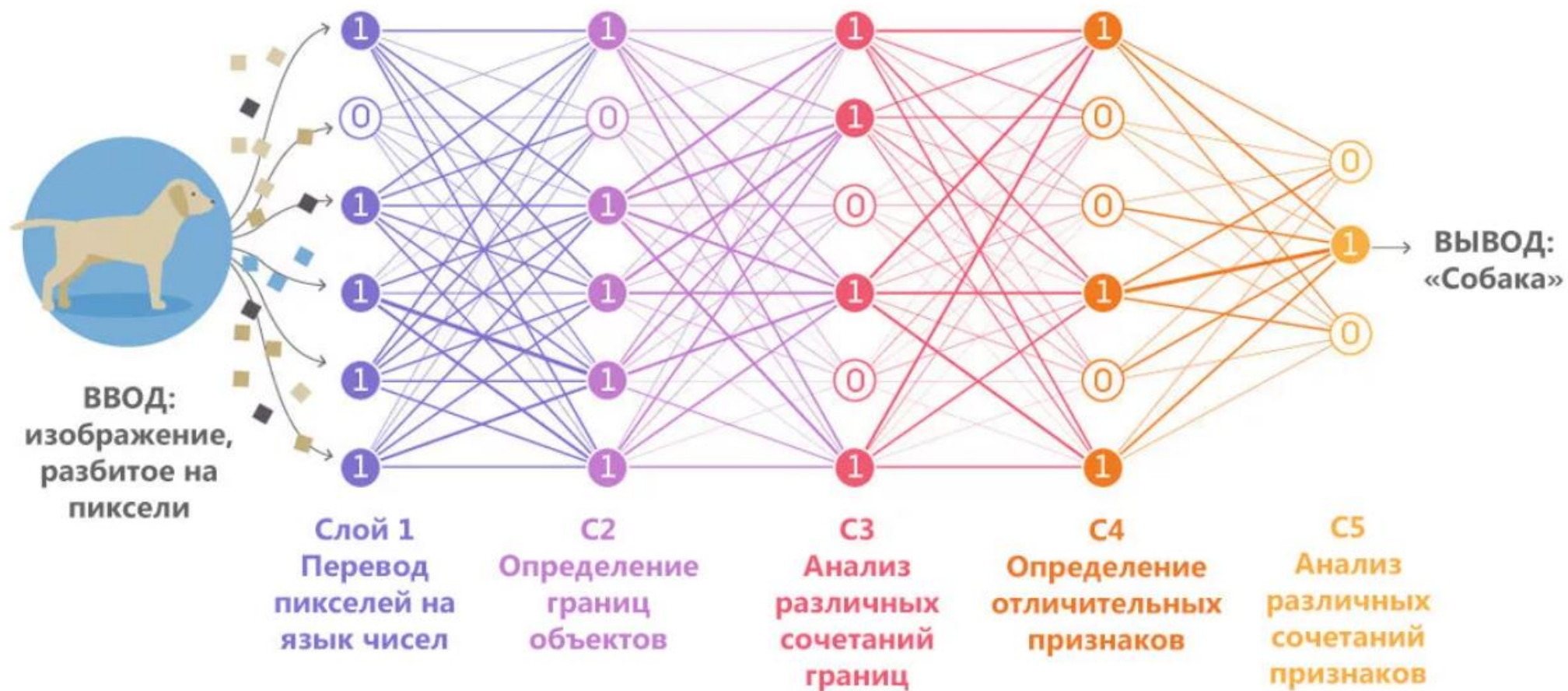
Каждый из объектов в выборке обладает сотней различных признаков, то основной трудностью будет графическое отображение такой выборки. Поэтому количество признаков уменьшают до двух или трёх, и становится возможным визуализировать их на плоскости или в 3D.

ГЛУБОКОЕ ОБУЧЕНИЕ

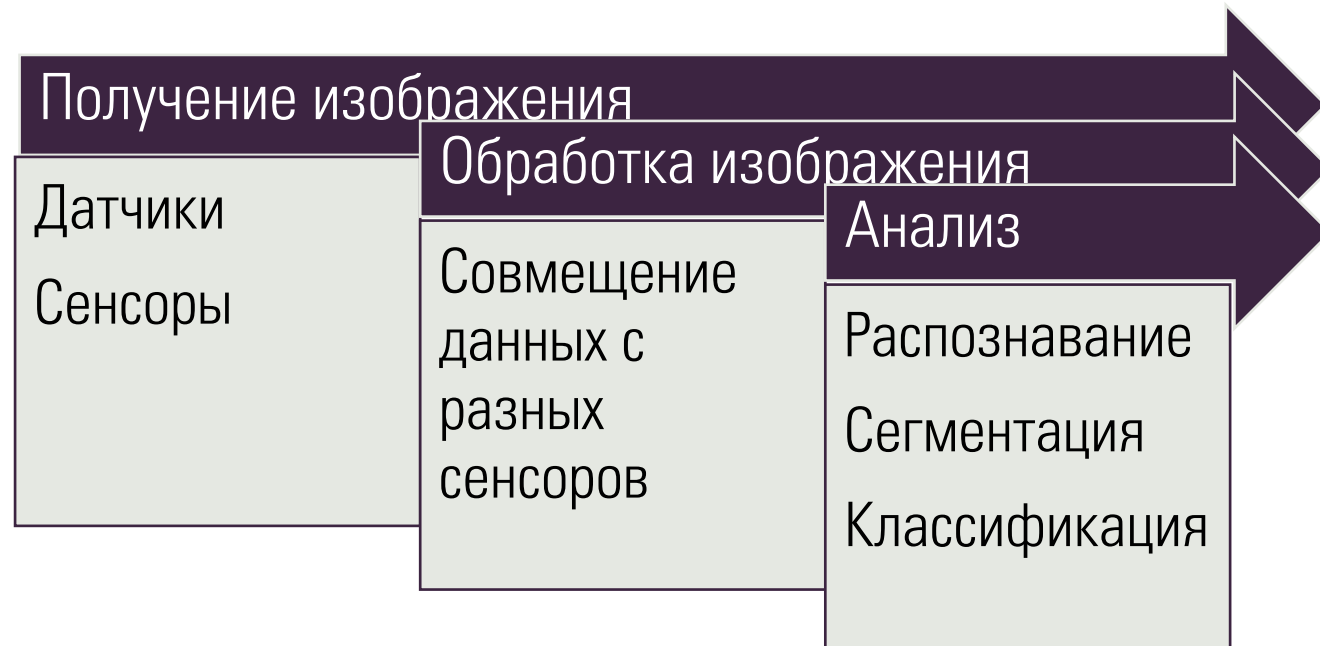
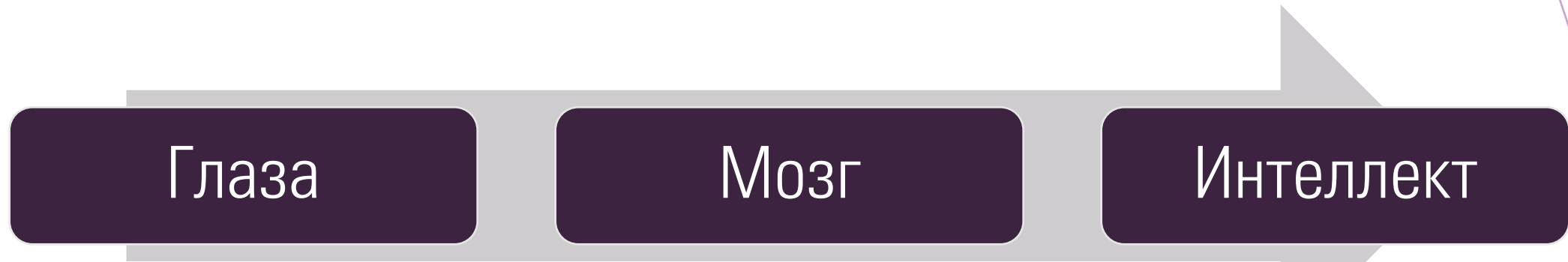
вид машинного обучения с использованием **многослойных нейронных сетей**, которые самообучаются на большом наборе данных: обработка изображений, машинный перевод, синтез и распознавание человеческой речи.



ГЛУБОКОЕ ОБУЧЕНИЕ

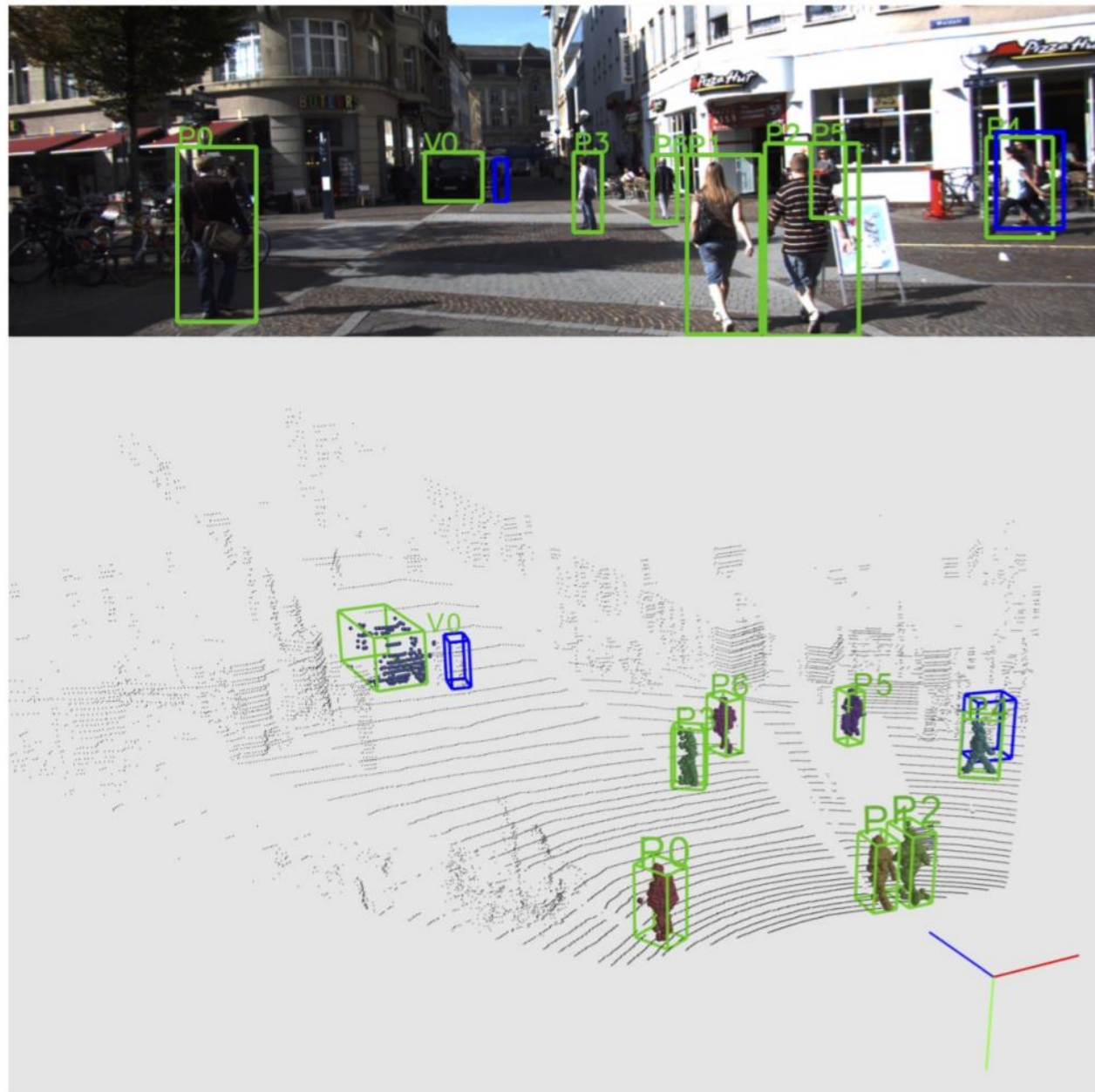


КОМПЬЮТЕРНОЕ ЗРЕНИЕ



ПРИМЕР

- Объемный куб правильного размера



АНАЛИЗ ДАННЫХ

Наука и практика анализа больших объемов данных с помощью всевозможных математических методов, в том числе машинного обучения, а также решение смежных задач, связанных со сбором, хранением и обработкой массивов данных.



ПРИМЕРЫ ПРОЕКТОВ

MAGENTA

- Создание музыкальных композиций
- Умеет создавать произведения искусства, основываясь уже на существующих.



DEEP PHOTO STYLE TRANSFER

- Копирует особенности стиля одной фотографии и переносит их на другие, при этом сохранив фотореалистичность.

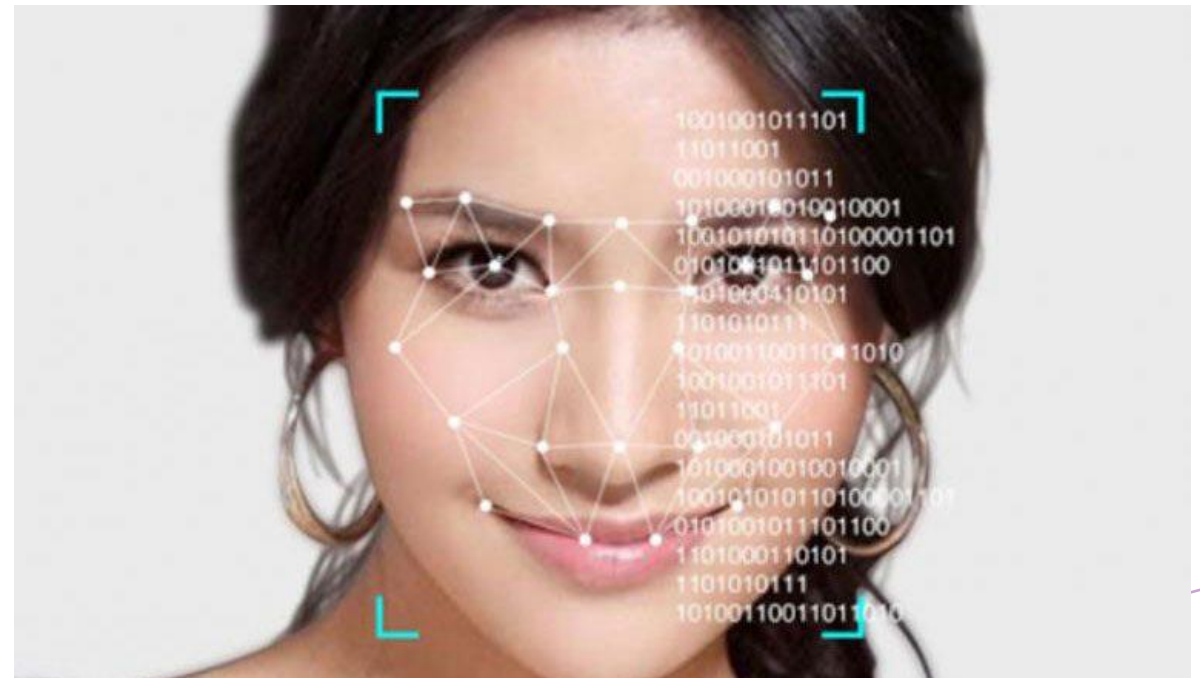


DEEP VOICE CONVERSION

- инструмент для работы со звуком: имитация голоса любого человека.
- для синтеза понадобятся реальные записи или живой пример.

FACE RECOGNITION

- распознавание лиц в реальном времени



БЛАГОДАРЮ ЗА ВНИМАНИЕ