

Курс по Машинному Обучению, Глубинному Обучению и Искусственному Интеллекту (ML, DL, AI)

Дзебоев Даниил

Общее введение в машинное обучение

Машинное обучение (ML), глубинное обучение (DL) и искусственный интеллект (AI) – это технологии, которые изменили наш подход к работе с данными, анализу и принятию решений. Эти методы позволяют машинам "учиться" на данных и адаптироваться к новым условиям, имитируя процессы мышления и решения задач.

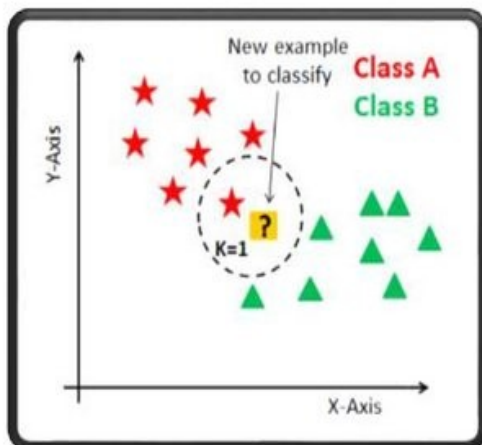


В современном мире ML и AI находят широкое применение: от анализа больших данных и предсказания трендов до компьютерного зрения и автоматизированного управления. Эти технологии становятся основой для создания сложных систем, способных решать задачи, которые ранее казались неподвластными автоматизации.

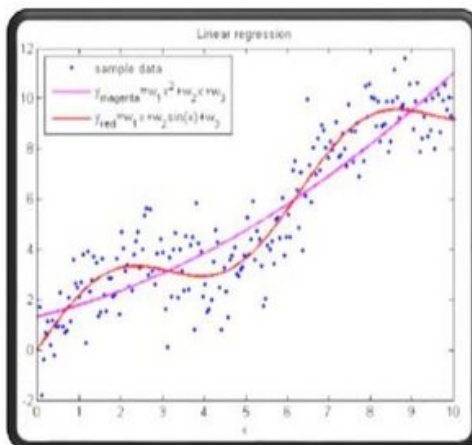


Виды машинного обучения

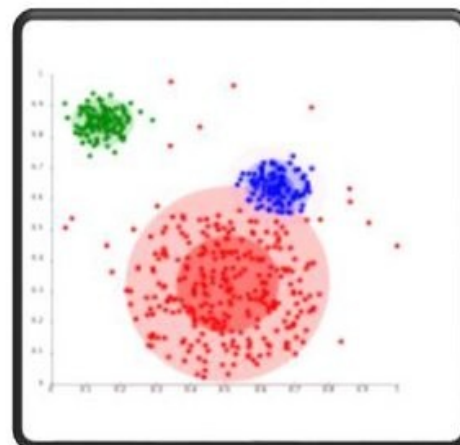
Классификация



Регрессия



Кластеризация



1. Классификация

Классификация – это задача машинного обучения, при которой модель учится распределять объекты по заранее определённым категориям или классам.

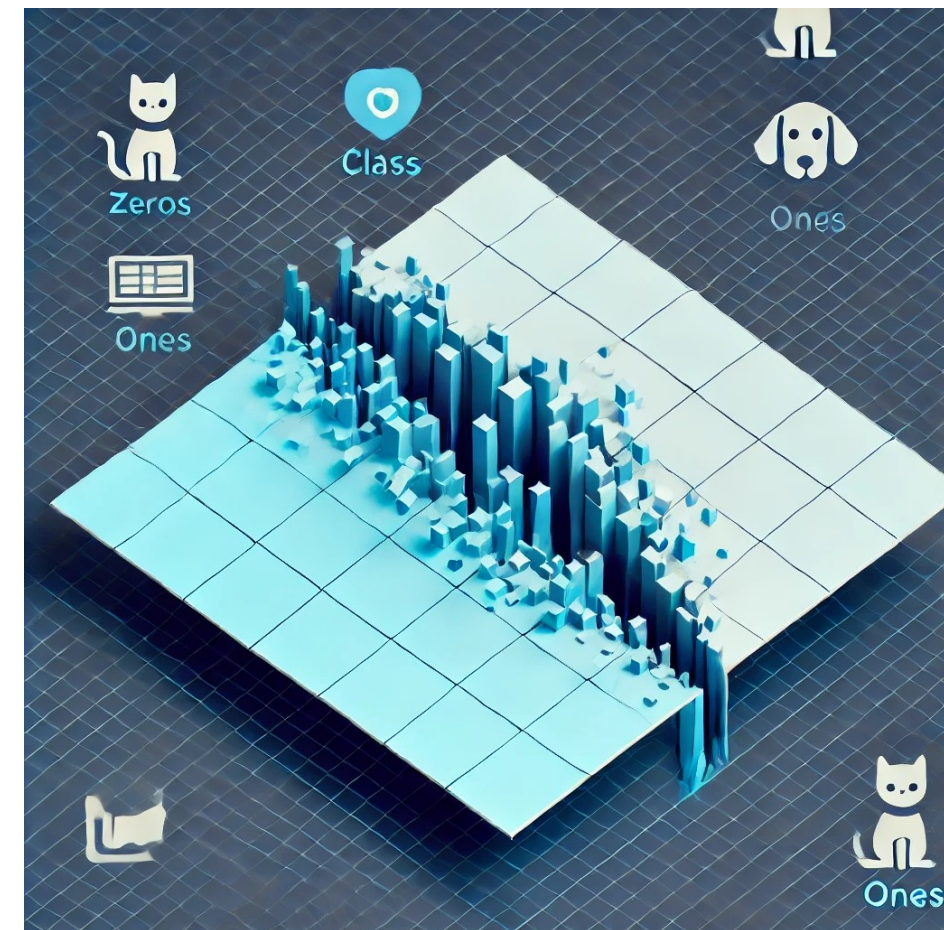
Примеры задач классификации:

Распознавание рукописных цифр (0–9).

Определение породы собаки по фотографии.

Классификация электронных писем на "спам" и "не спам".

Модели классификации возвращают дискретные значения, такие как классы или метки, например, "Да/Нет", "Кошка/Собака".



2. Регрессия

Регрессия используется для предсказания непрерывных значений. Она помогает определить зависимость между переменными и прогнозировать будущие значения.

Примеры задач регрессии:

Прогнозирование цен на жильё.

Предсказание температуры на завтра.

Определение зависимости продаж от затрат на рекламу.

В регрессии модель выдаёт непрерывное числовое значение, например, цену, процент или количество.



3. Кластеризация

Кластеризация относится к методам обучения без учителя, при котором данные группируются на основе схожих признаков. Модель сама определяет, как сгруппировать данные без предварительных меток.

Примеры задач кластеризации:

Сегментация клиентов на группы по их поведению.

Группировка изображений по схожести.

Разделение генов в биологической науке.

Кластеризация помогает находить скрытые закономерности в данных и группировать объекты по схожим характеристикам.

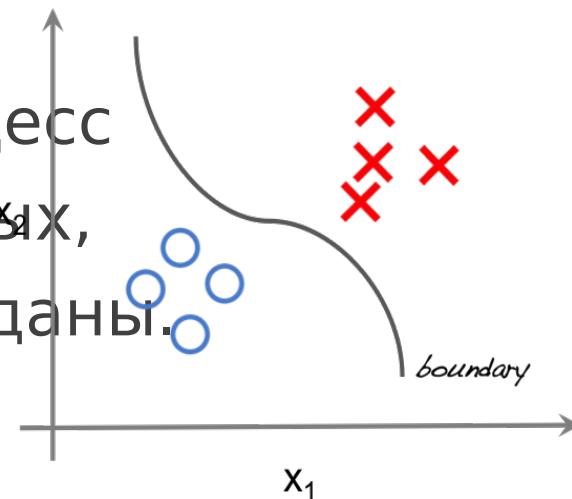


Концепция обучения с учителем и без

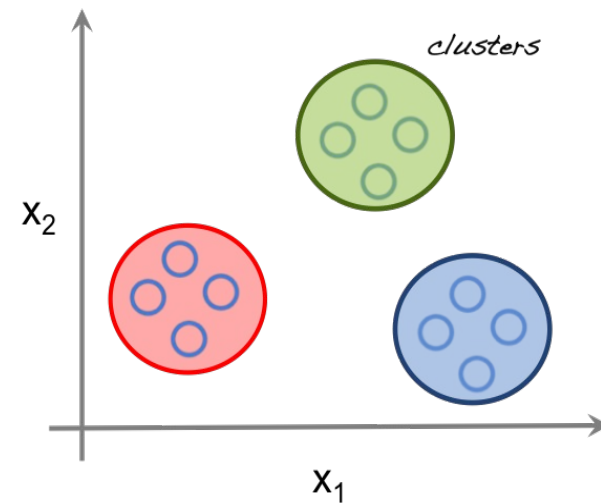
Обучение с учителем - это когда у нас есть большой накопленный опыт и мы хотим научить машину переводить одну информацию в другую. Копировать алгоритм с улавливанием внутренних связей на основе больших данных, прогнозировать результат. Мы её в каком-то смысле сами настраиваем на определённый результат

Supervised learning

Обучение без учителя - это процесс поиска закономерностей в данных, в которых они изначально не заданы.



Unsupervised learning



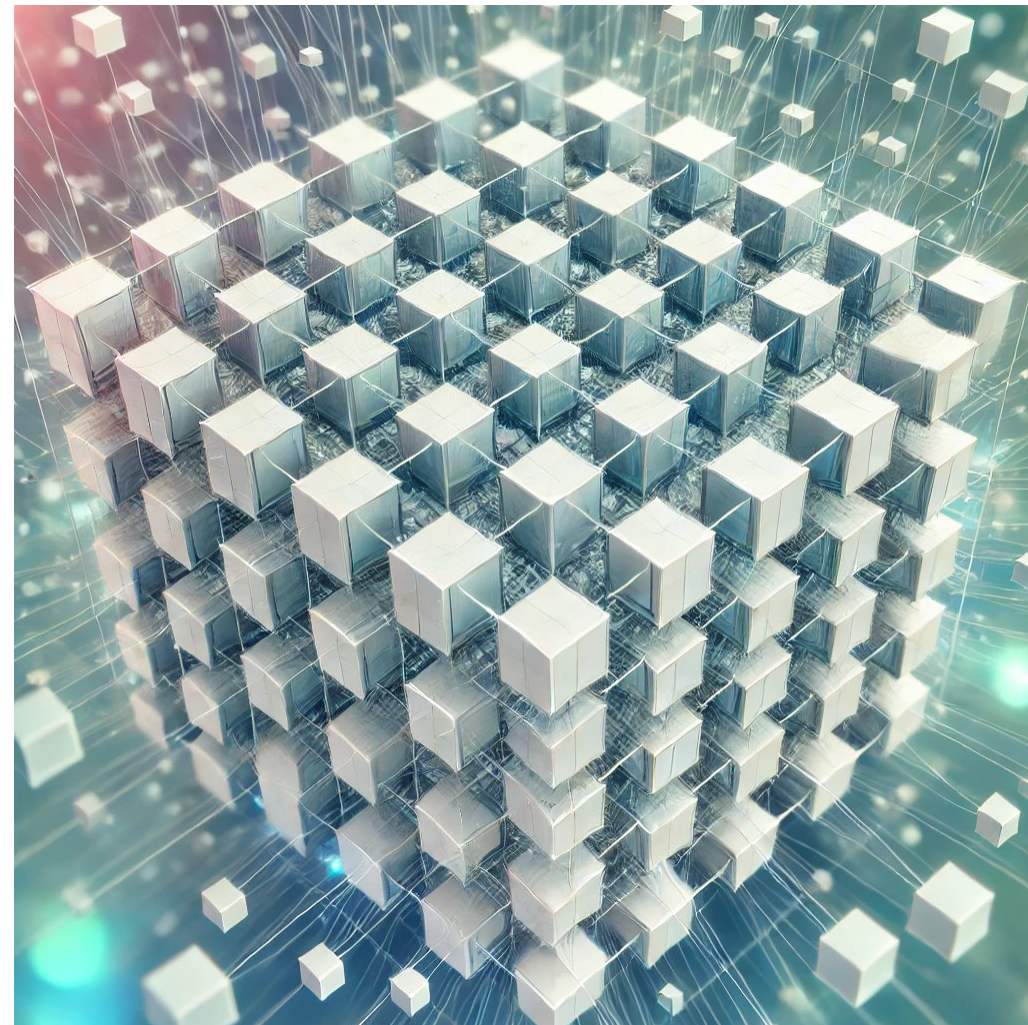
Виды организации данных

Машинное обучение работает с различными форматами данных, и выбор правильной структуры данных играет важную роль в эффективности модели. Рассмотрим основные виды организации данных:

1. Многомерные массивы

Многомерные массивы (тензоры) представляют собой данные, организованные в несколько измерений. В машинном обучении часто используются массивы данных для хранения изображений, звуковых сигналов или временных рядов.

Пример: трёхмерный массив для хранения изображений (ширина, высота, цветовые каналы)



2. Векторы

Вектор – это одномерный массив, содержащий числовые значения. Векторы часто используются для представления признаков объектов (например, набор характеристик).

Пример: вектор признаков, описывающий объект по его атрибутам (размер, цвет, вес).

$$\begin{pmatrix} \text{с} \\ \text{т} \\ \text{о} \\ \text{л} \\ \text{б} \\ \text{е} \\ \text{ц} \end{pmatrix}^t = (\text{с} \text{ т} \text{ р} \text{ о} \text{ ч} \text{ к} \text{ а})$$

3. Таблицы

Табличные данные состоят из строк и столбцов, где каждая строка представляет объект, а каждый столбец – его признак. Таблицы – это распространённая форма организации данных, особенно в бизнесе и статистике.

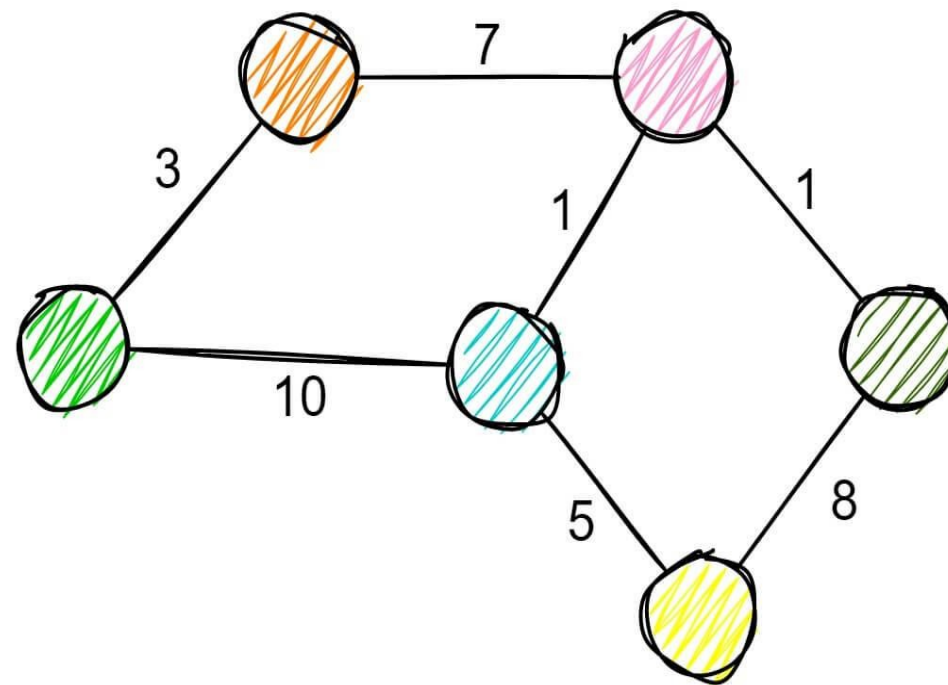
Пример: таблица с данными о клиентах, где каждый столбец – это один из признаков клиента (возраст, пол, доход).



4. Графы

Графы – это структура, состоящая из узлов и рёбер, которые связывают эти узлы. Графы особенно полезны для моделирования связей между объектами.

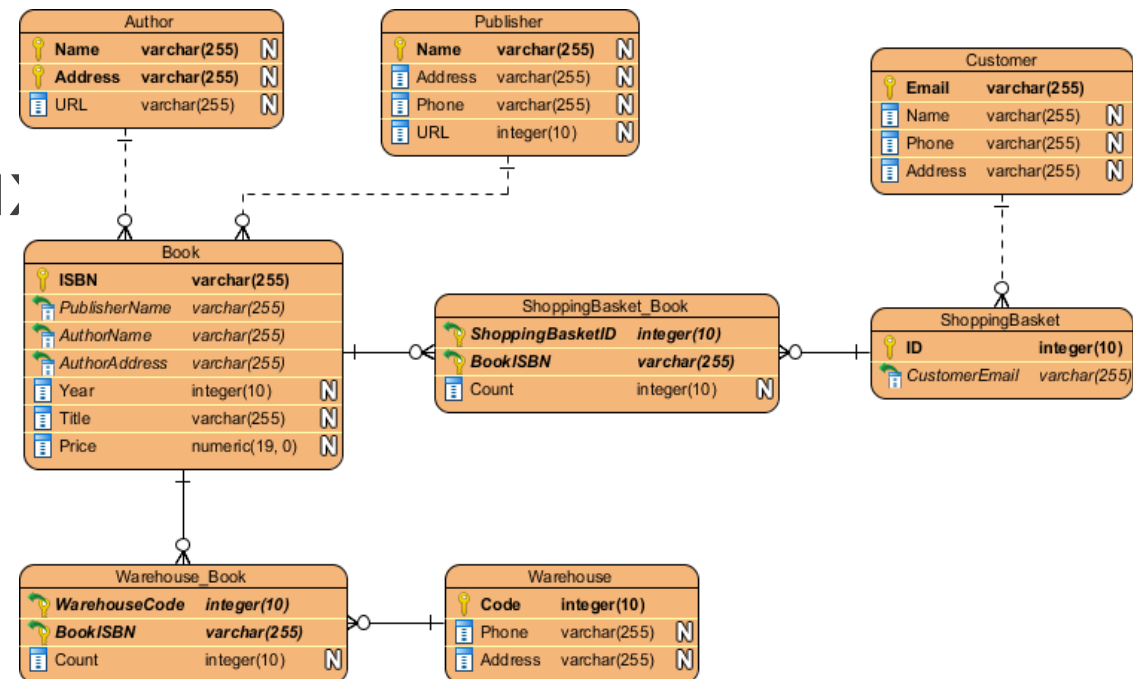
Пример: социальные сети, где узлы – это пользователи, а рёбра – их взаимодействия.



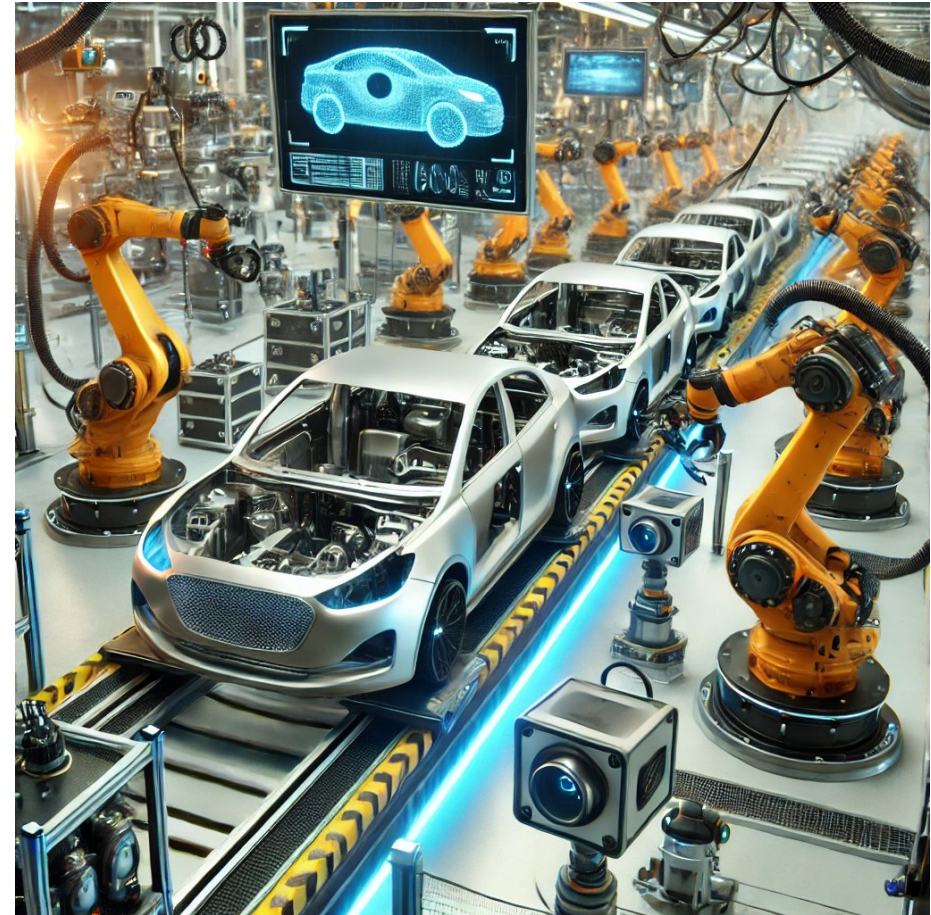
5. Реляционные данные

Реляционные данные организованы в виде связанных между собой таблиц, которые можно объединять по ключам. Это основа для реляционных баз данных, где данные представлены в виде взаимосвязанных сущностей.

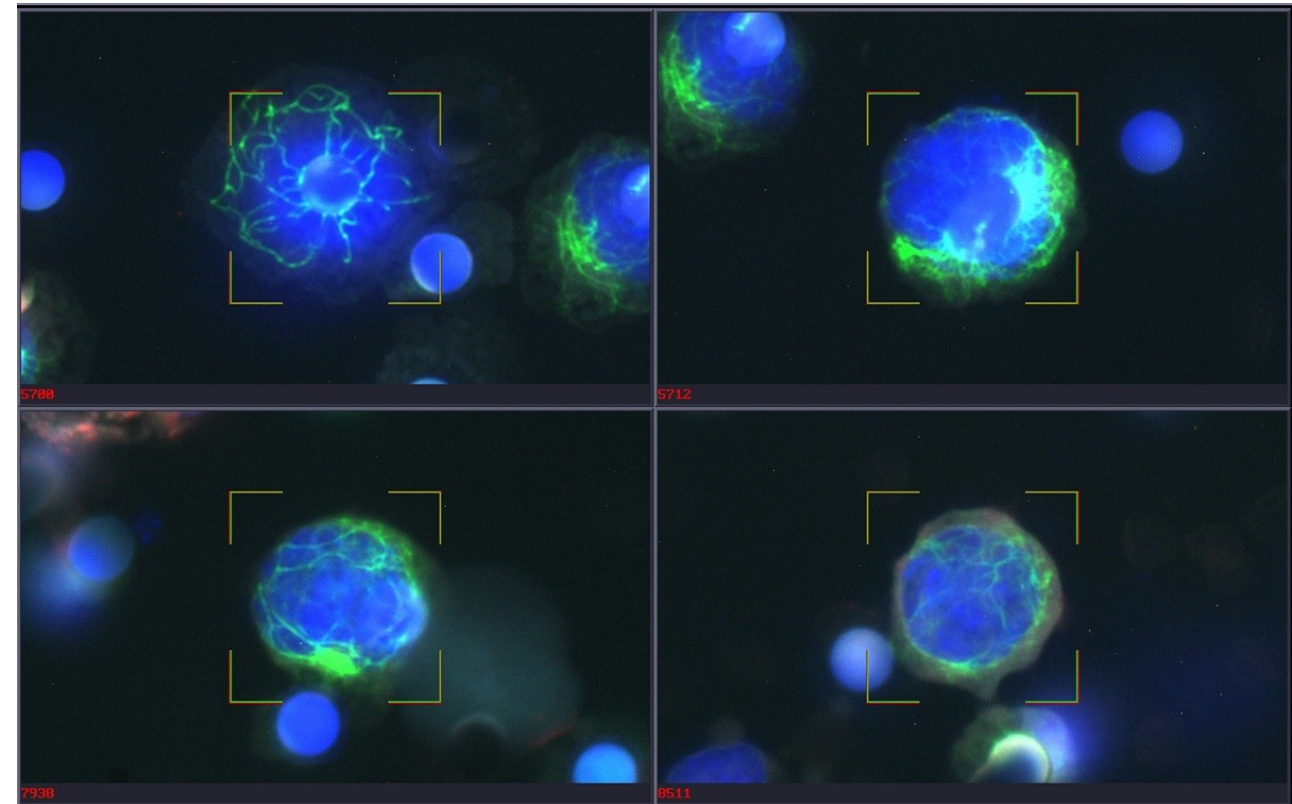
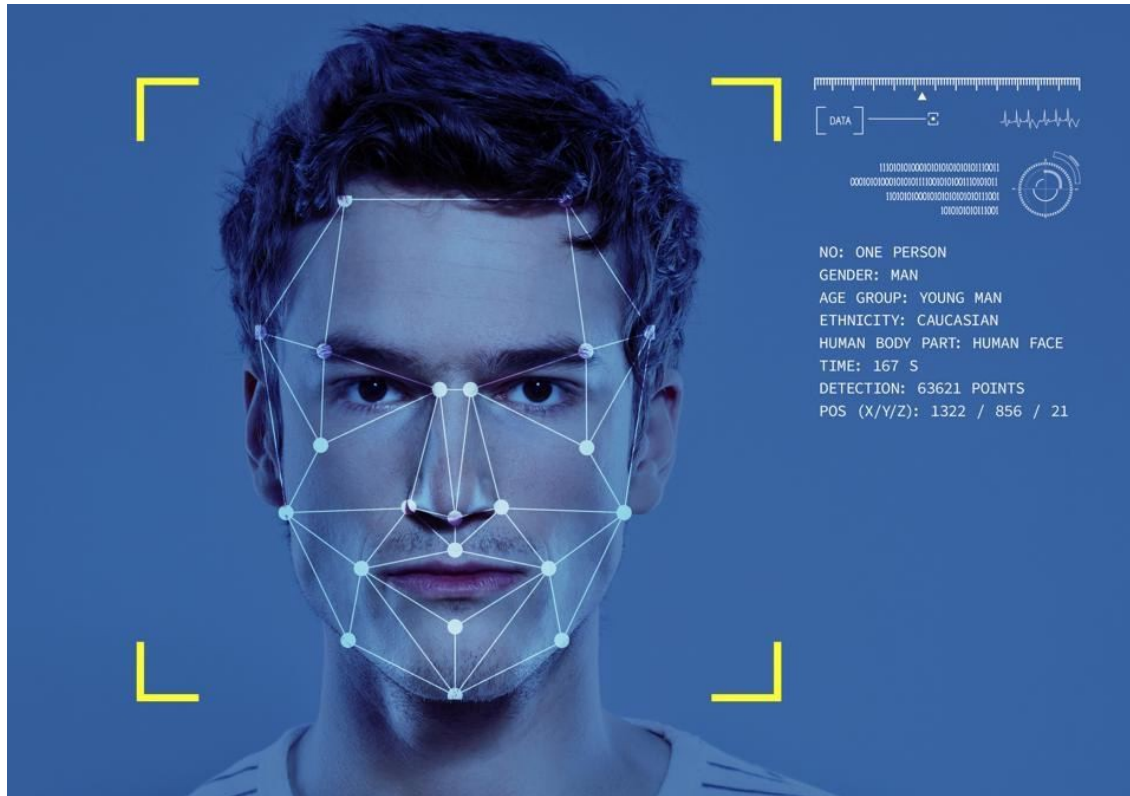
Пример: база данных, которая связывает заказы с клиентами и продуктами через уникальные идентификаторы.



Применение машинного обучения в промышленности



Применение машинного обучения в биометрии



Системы прогнозирования трендов



Системы поиска и структуризации неструктурированной информации

Очень важно в условиях постоянного притока эксабайтов данных со временем не погрязнуть в них и не потерять их. Для оперативного поиска как в данных, так и в самом социуме нужны алгоритмы структуризации данных и выделение признаков.

В случае данных это к примеру группировка гигабайтов фотографий внутри телефона на людей, места, «вайбы»

В случае социума это поиск соответствий в экономике или людей на сайтах знакомств. Я активно занимался обеими направлениями.

И вообще очень заинтересован в оцифровке вайбов,

Таким образом, искусственный интеллект - это мощнейшая сфера современной научной мысли. Он не может идти в одиночку. Он обязан идти вкуче с автоматизированными системами управления предприятием. С роботизированными руками, автоматизированными системами. Но искусственный интеллект имеет применения в управлении роботизированной техникой и в сохранении человеческого сознания, истории, структуризации данных

Спасибо за внимание!