

Предсказание магнитных свойств наночастиц для биомедицинских применений

Машинное обучение



Зачем нам машинное обучение?

Есть зависимость целевого параметра от множества факторов

Расчетные методы неэффективны для таких сложных систем

Экспериментальный перебор большого числа факторов – задача неэффективная и неблагодарная

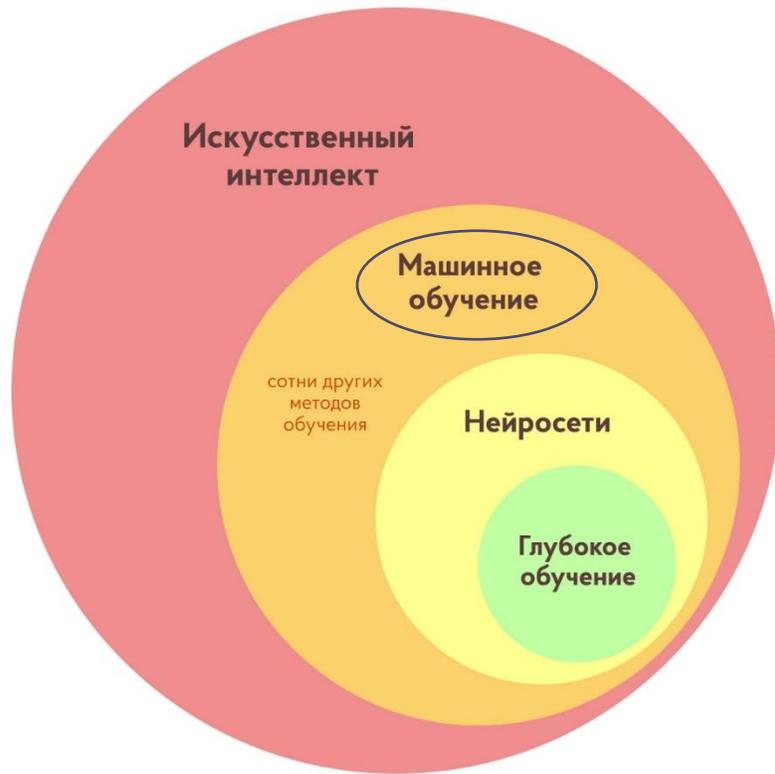
Встает вопрос: а можно ли научиться оценивать целевые параметры ещё до проведения эксперимента, чтобы понимать, частицу с примерно какими свойствами нам нужно получить для того, чтобы она была эффективна для биомедицинских применений?

Ответ - да

Зачем нам машинное обучение?

Существуют неявные зависимости между параметрами частицы, условиями эксперимента и значениями SAR, r_1/r_2 релаксаций

Методы машинного обучения позволяют построить функцию, которая отражает эти зависимости



Наука о данных — это междисциплинарная область, которая использует научные методы, процессы, алгоритмы и системы для извлечения знаний и идей из зашумленных, структурированных и неструктурированных данных, а также для применения знаний и практических идей из данных в широком спектре прикладных областей

Искусственный интеллект (ИИ) — это интеллект, демонстрируемый машинами, в отличие от естественного интеллекта, демонстрируемого животными, включая людей. Ведущие учебники ИИ определяют эту область как изучение интеллектуальных агентов – любая система, которая воспринимает свое окружение и предпринимает действия, максимизирующие его шанс на достижение своих целей"

Машинное обучение (ML) – это изучение компьютерных алгоритмов, которые могут автоматически улучшаться с помощью опыта и использования данных

Pipeline проекта по машинному обучению

Сбор и обработка данных (очень важный этап): **garbage in – garbage out**

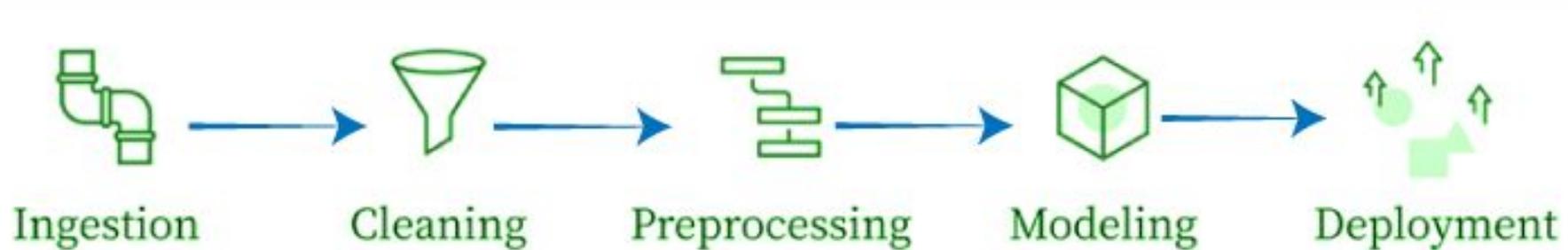
Описание используемого датасета

Выбор используемого алгоритма: tree-based algorithms, линейная регрессия, k-NN, бустинг и тд.
Обучение и тест модели МО. Кросс-валидация. Функции потерь и метрики качества моделей

Интерпретация моделей МО (по возможности)

Валидация моделей на ранее не виденных образцах

Определение границ применимости полученной модели



Инструменты

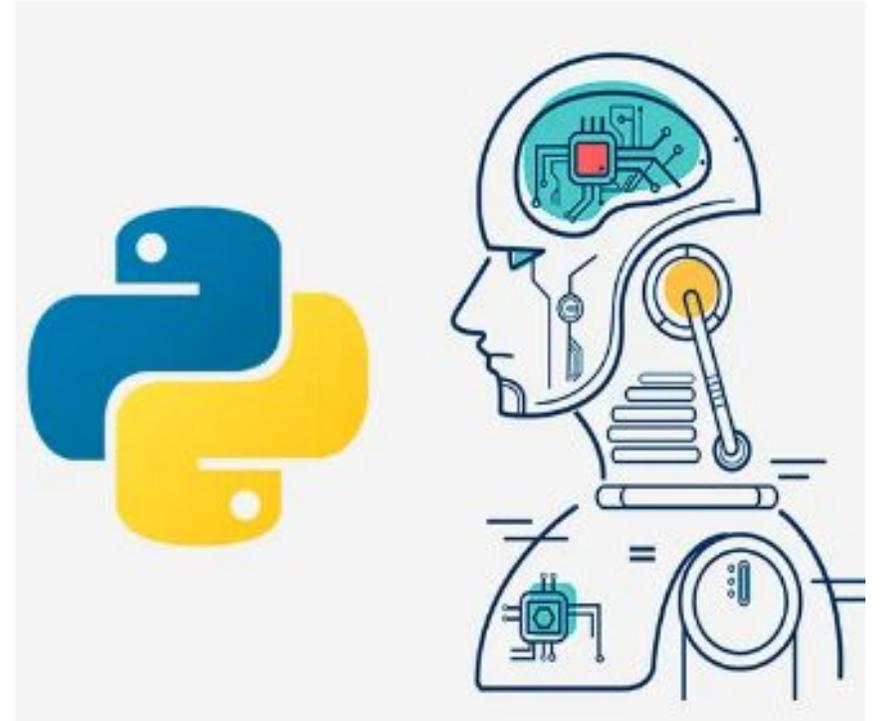
Будем использовать язык программирования Python

Python — это один из самых распространённых языков программирования

Существует огромное количество пакетов, которые позволяют решать с помощью этого языка самые разные задачи

Можно наугад назвать какую-нибудь сферу деятельности и в экосистеме Python, почти гарантированно, найдутся отличные инструменты для решения специфических задач из этой сферы

В сфере машинного обучения Python показывает себя наилучшим образом



Pure python

Типы данных: integer, float, string, list, dictionary, tuple и т.д. и их методы

Условия: if, elif, else

Циклы: for, while

Библиотеки

NumPy: поддержка многомерных массивов, то есть матрицами; поддержка высокоуровневых математических функций

Pandas: надстройка на NumPy - специальные структуры данных и операции для манипулирования числовыми таблицами. DataFrame – таблица с колонками

Matplotlib: визуализация данных двумерной и трёхмерной графикой

Scikit-learn: библиотека, предназначенная для машинного обучения

Остальные: rdkit, lazypredict, optuna