*Задание (последовательность указанных действий определяется вами, обратите внимание на важность этого момента):*

* Разработать дескрипторы материалов на основании имеющихся в базе данных (pymatgen и другие в помощь), обосновать их интерпретируемость с позиций здравого смысла и показать на feature importance
* Нормализовать данные и подготовить для построения алгоритма машинного обучения
* Выбрать лучший алгоритм МО на основании метрик точности и ошибок и !5-fold cross-validation

**! Спойлер! Алгоритм из статьи не лучший**

* Сделать оптимизацию гиперпараметров моделей МО
* Показать интерпретируемость модели

*Предсказываемые свойства:* Km, Kcat

*Результат:*

Рабочий код, выполняющий все эти функции и выводящий читаемые и понятные графики. Чем лаконичнее, читаемее и быстрее – тем лучше. Отдельный плюсик тем, кто напишет код, запускающийся у других людей без внесения каких-либо корректировок.

*По всем вопросам можно обращаться к Никите Серову, Юле Разливиной и Оле Капустиной.*

*Данные:*

Это данные собранные из статей по нанозимам с пероксидазной активностью. Они включают в себя данные о материале (состав, размеры, структура, условия синтеза и анализа) и его каталитической активности по кинетике Михаэлиса.

Обозначения параметров:

formula - химическая формула

Km - константа Михаэлис-Ментен, mM

Kcat - константа каталитической активности, равная Vmax/Ccat, s-1

Syngony - кристаллическая система образца, категориальный признак (0-аморф, 1-7 системы)

length - длина наноматериала, нм

width - ширина наноматериала, нм

depth высота наноматериала, нм

pol - категориальный признак, показывающий наличие полимера в синтезе (0-нет, 1-да)

surf - категориальный признак, показывающий наличие нейтральных ПАВ в синтезе (0-нет, 1-да)

Subtype - субстрат, на котором проводили измерение каталитической активности

ph - ph буфера, в котором измеряли каталитическую активность

temp - температура, при которой измеряли каталитическую активность

Cper - концентрация H2O2 (mM), при которой измеряли каталитическую активность

Csub - концентрация хромогенного субстрата (mM), при которой измеряли каталитическую активность

Ccat - концентрация наночастиц (mkg/ml), при которой измеряли каталитическую активность