# Поправки в первое задание

А. Для финализации этапа предсказания химии, необходимо написать выводы в абзац по предобработке данных.

Все этапы указанные в задании, для целевых переменных.

«EDA – выводы по качеству данных:

1. Пропуски
2. Нули
3. Дисперсия
4. Выбросы»

Б. Провести предсказание, НЕ ЗАГЛЯДЫВАЯ в будущее, т.е. данные «химсталь\_последний» и «химшлак\_последний» некорректно(нельзя!) использовать в фичах. Эти данные появляются на последнем этапе обработки, значит ТОЛЬКО вместе с целевыми переменными, без целевых переменных мы этих данных знать не можем.

Далее, проводя предсказания, результат для всех использованных моделей, вывести используя метрики R2, MAE, MSE, MAPe (средняя относительная ошибка **в процентах**) в таблице.

Можно например в такой:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Модель | R2 | MAE | MSE | MAPe | Лучшие параметры модели |
| Linreg… |  |  |  |  |  |
| Gradientboosting |  |  |  |  |  |
| ….. |  |  |  |  |  |

Коротко напишите - какой метрикой по вашему мнению следует руководствоваться при выборе модели в первую очередь и почему?

(*Добрый совет: постройте графики предсказаний и фактических значений, какие различия видите? Метрика, что-то подчеркивает или нивелирует, любая. Что тут стоит подчеркнуть?*)

Вывод – какая модель в вашей табличке по вашему лучше?

Для второго этапа результаты также попрошу представить в аналогичном виде.