Задание для курсового проекта

Метрика:

R2 - коэффициент детерминации (sklearn.metrics.r2\_score)

Сдача проекта:

1. Прислать в раздел Задания Урока 10 ("Вебинар. Консультация по итоговому проекту")

ссылку на ноутбук в github или public kaggle notebook.

2. Необходимо получить R2 > 0.6.

Примечание:

Все файлы csv должны содержать названия полей (header - то есть "шапку"),

разделитель - запятая. В файлах не должны содержаться индексы из датафрейма.

\_\_\_\_\_\_\_\_\_\_\_\_

Рекомендации для файла с кодом (ipynb):

1. Файл должен содержать заголовки и комментарии

2. Повторяющиеся операции лучше оформлять в виде функций

3. По возможности добавлять графики, описывающие данные (около 3-5)

4. Добавлять только лучшую модель, то есть не включать в код все варианты решения проекта

5. Скрипт проекта должен отрабатывать от начала и до конца (от загрузки данных до выгрузки предсказаний)

6. Весь проект должен быть в одном скрипте (файл ipynb).

7. При использовании статистик (среднее, медиана и т.д.) в качестве признаков,

лучше считать их на трейне, и потом на валидационных и тестовых данных не считать

статистики заново, а брать их с трейна.

8. Проект должен полностью отрабатывать за разумное время (не больше 10 минут),

поэтому в финальный вариант лучше не включать GridSearch с перебором

большого количества сочетаний параметров.