

Добрый день, Анатолий!

Практически по всем пунктам критериев Вы отлично выполнили задание, замечательная и красивая работа! :)

Требования к оформлению:

- + на github: отличный подробный readme-файл, содержащий цели и задачи, описание данных и этапов работы над проектом;
- + структура оформления ноутбука: всё последовательно разделено на логические части, описаны цели и задачи, каждый этап работы, подробные сопровождающие пояснения и содержательные выводы о проделанной работе;
- + общие правила оформления: читаемый понятный код с грамотными названиями функций и переменных, графики построены по всем правилам визуализации, соблюдение стандартов PEP-8, отформатированные выводы в отдельных ячейках типа Markdown.

Анализ и обработка данных:

Отлично оформили этап первичного исследования данных: здорово, что посмотрели на число пропусков, дубликатов и выбросов; и грамотно их обработали. В итоге получился хороший датасет с достаточным количеством признаков. Можете также попробовать для удобства использовать метод describe.

Глубокий подробный разведывательный анализ; хорошо, что делаете выводы, отдельный плюс за информативные визуализации, особенно карту корреляций.

Плюс за использование статистических методов и методов отбора признаков.

В целом, достаточный и глубокий анализ, нечего добавить.

Применение ML и DL:

Здорово, что последовательно “по всем правилам” протестировали несколько известных методов. Зачастую случается так, что классические алгоритмы для некоторых задач показывают гораздо лучшие метрики 😊 Замечательные выводы по ходу выполнения работы, нечего добавить.

CatBoost - сейчас самый популярный алгоритм, дающий, как правило, самые лучшие метрики, про него, кстати, часто спрашивают на собеседованиях - очень хорошо, что разобрались с ним.

Здорово, что пытались подобрать гиперпараметры при помощи GridSearchCV - чтобы “дождаться” результатов можете или уменьшить число и диапазон подбираемых гиперпараметров, или воспользоваться библиотекой optuna - она работает эффективнее за счёт того, что не обрабатывает те значения из предложенного диапазона, которые заведомо являются неоптимальным направлением.

Из советов и пожеланий:

Можете также посмотреть про модуль tqdm - с его помощью легко следить за прогрессом выполнения операций в цикле, обучения моделей и применения apply. Ради интереса можете также потестировать нейросети:)

Спасибо за выполненное задание! Отзыв подготовила ментор Мария Жарова. Если возникнут вопросы, можете обратиться ко мне в канал 08_final в пачке, постараюсь

ответить на все вопросы и разобраться с моментами, которые вызывают трудности.
Удачи в обучении!