

Добрый день, Станислав!

Практически по всем пунктам критериев Вы отлично справились с заданием, замечательная и красивая работа! :)

Требования к оформлению:

- + 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;
- + 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 наличие пропусков, дубликатов и выбросов.

Здорово, что построили графики распределений признаков, упростить этот процесс может использование функции `pairplot`. Плюс за карту корреляций признаков и тестирование статистических гипотез. И отдельный плюс за использование `SweetViz`!

Здорово, что также сопровождали работу подробными комментариями в отдельных ячейках `markdown`.

Предобработка данных во `feature engineering` замечательная: здорово, что оформили код для обработки каждого признака в функции, которые можно переиспользовать. Именно так обычно и принято делать в настоящих проектах:)

Отдельный плюс за тестирование статистических гипотез и использование методов отбора признаков.

Качественно сделали очистку данных и непосредственную подготовку их на вход модели, здесь нечего добавить.

Применение ML и DL:

Здорово, что последовательно “по всем правилам” протестировали все известные методы от простых к сложным. Зачастую случается так, что классические алгоритмы для некоторых задач показывают гораздо лучшие метрики - но, как видим, не всегда:) `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 `RandomizedSearchCV` и, в особенности, `optuna` - это одна из самых распространённых современных библиотек для подбора гиперпараметров (она работает эффективнее за счёт того, что не обрабатывает те значения из предложенного диапазона, которые заведомо находятся

на неоптимальном направлении). Если интересно, можете ещё познакомиться с библиотекой `hugorpt`, она тоже сейчас довольно популярна.

Отдельно хочется отметить:

Хорошо, что пояснили код комментариями по всем стандартам PEP-8 и сопроводили всё решение подробными пояснениями в Markdown-ячейках.

Отличная реализация варианта решения в продакшене в `docker`, всё работает верно и без перебоев.

Серьёзно подошли к выполнению проекта, это очень здорово! Ничего не упустили, замечательная работа 👍 Можно гордиться таким качественно выполненным реп-проектом 🔥

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`.

Спасибо за выполненное задание! Если возникнут вопросы, можете обратиться в канал `08_final` в пачке, постараюсь ответить на все вопросы и разобраться с моментами, которые вызывают трудности. Удачи в обучении!