

Тестовое задание. Пояснения

EDA_utils.py - модуль с функциями для эксплоративного анализа

TrainModel.py - модуль подбора параметров модели, загрузки тренировочных данных, сохранения модели

Preprocessor.py - модуль (заготовка) предобработки данных

Predictor.py - модуль для предсказаний на новых данных

Jupyter notebook **Using_EDA_utils_example.ipynb** демонстрирует применение модуля EDA_utils.py

Jupyter notebook **Using_TrainModel_and_Predictor_example.ipynb** демонстрирует обучение модели, валидацию и предсказания на новых данных

Если применять подобный инструмент в реальной ситуации, считаю важным сделать как минимум следующие доработки:

- заменить все выводы на логирование
- добавить возможность вызова и подбора параметров различных алгоритмов (не только knn)
- добавить отбор признаков
- для рассмотренной задачи бинарной классификации проанализировать, какой из классов важнее хорошо прогнозировать (или классы равнозначны); в зависимости от этого выбрать наиболее подходящую метрику
- засекаать время обучения модели и время предсказания