

Проект Docker_Flask_Pi pipeline

Подтема

Итоговый проект по курсу «Машинное обучение в бизнесе»

Задача: предсказать по описанию клиента банка вероятность его ухода.





Дмитрий Яковлев

Закончил факультет искусственного интеллекта
GeekBrains..

Закончил курсы повышения квалификации в МГТУ им.
Баумана.

Немного о себе.

Инженер по работе с проектировщиками и клиентами на
предприятии, производящем высоковольтное
оборудование.



План проекта





Решение задачи / План работы

- Исследование данных на платформе Kaggle.
- Исследование рабочего датасета: поиск аномалий, анализ распределений, корреляций, исследование методов работы с количественными и категориальными признаками, исследование выбросов, поиск оптимального алгоритма построения регрессора, определение полезных признаков.
- Генерация новых признаков, исследование их корреляции.
- Разработка пайплайна на базе классификатора.
- Разработка product ready приложения на основе фреймворка Flask, обернутого в Docker.



Выполненная работа

Сам выбранный мной самостоятельно датасет на Kaggle по ссылке <https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers> ни чего интересного как оказалось из себя не представляет. Обычный кagловский причесанный и прилизанный датасет, на котором легко выбил метрики, близкие к единице.

Выбрал для простоты из датасета 6 самых значимых признаков, создал из них еще четыре признака.

Задачей проекта было предусмотрено свести всю работу решателя к единому пайплайну на основе соответствующего метода библиотеки Sklearn. Поэтому были созданы на основе этого метода соответствующие классы для предобработки, генерации новых признаков и классификации. И все эти классы были сведены в один пайплайн. Создал для проверки приложения два ноутбука (Step_1 и Step_2).

Далее на основе этого пайплайна было создано на основе фреймворка Flask приложение и это приложение было обернуто в Docker и выложено на Github.